

This is a repository copy of *Part by part : synthetic biology parts used in solventogenic Clostridia*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/125408/>

Version: Accepted Version

Article:

Gyulev, Ivan, Willson, Benjamin James, Hennessy, Rosanna Catherine et al. (3 more authors) (2018) *Part by part : synthetic biology parts used in solventogenic Clostridia*. ACS Synthetic Biology. pp. 311-327. ISSN 2161-5063

<https://doi.org/10.1021/acssynbio.7b00327>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Part by part: synthetic biology parts used in solventogenic Clostridia

Authors:

Ivan S. Gyulev^{1*}, Benjamin J. Willson^{1*}, Rosanna C. Hennessy², Preben Krabben³, Elizabeth R. Jenkinson⁴, Gavin H. Thomas^{1§}

Author affiliation:

1. Department of Biology, University of York, Wentworth Way, York, YO10 5DD, United Kingdom

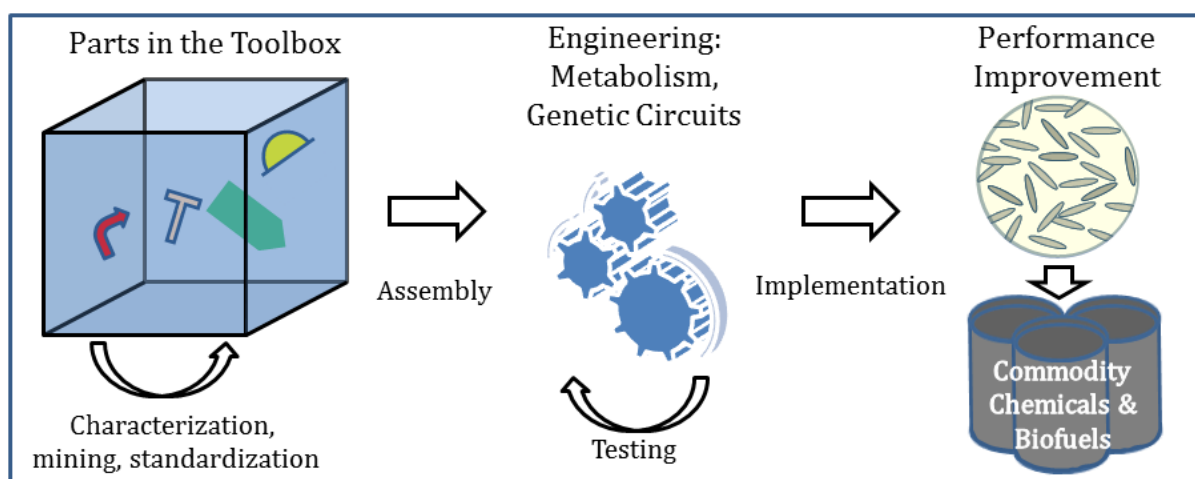
2. Microbial Ecology and Biotechnology, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

3. 9 Marsh Lane, Didcot, Oxfordshire, OX11 8FD, United Kingdom

4. Green Biologics Limited, Milton Park, Abingdon, Oxfordshire, OX14 4RU, United Kingdom

*** These authors contributed equally to this work.**

§ - Phone: 44-1904-328678. E-mail: gavin.thomas@york.ac.uk.



Graphical Table of Contents.

ABSTRACT: The solventogenic *Clostridia* are of interest to the chemical industry because of their natural ability to produce chemicals such as butanol, acetone and ethanol from diverse feedstocks. Their use as whole cell factories presents multiple metabolic engineering targets that could lead to improved sustainability and profitability of *Clostridium* industrial processes. However, engineering efforts have been held back by the scarcity of genetic and synthetic biology tools. Over the last decade, genetic tools to enable transformation and chromosomal modifications have been developed, but the lack of a broad palette of synthetic biology parts remains one of the last obstacles to the rapid engineered improvement of these species for bioproduction. We have systematically reviewed existing parts that have been used in the modification of solventogenic *Clostridia*, revealing a narrow range of empirically chosen and non-engineered parts that are in current use. The analysis uncovers elements, such as promoters, transcriptional terminators and ribosome binding sites where increased fundamental knowledge is needed for their reliable use in different applications. Together, the review provides the most comprehensive list of parts used and also presents areas where an improved toolbox is needed for full exploitation of these industrially important bacteria.

Keywords

Clostridium, ABE fermentation, solventogenesis, transcription termination, reporters, promoters

Abbreviations

ABE: acetone-butanol-ethanol

ACE: allele-coupled exchange

aSD: anti Shine-Dalgarno sequence, found at the 3' end of 16S rRNA

CDS: coding sequence

FbFP: flavin-binding fluorescent protein

RBS: ribosome binding site

RNAP: RNA-polymerase

SD: Shine-Dalgarno sequence

UTR: untranslated region

Acknowledgements:

We acknowledge the BBSRC for support for Ivan Gyulev through an iCASE PhD studentship with Green Biologics Ltd (BB/M016331/1), Benjamin Willson from an IB Catalyst grant for Project DETOX (BB/N01040X/1) and Rosanna Hennessy for a BBSRC IUK award (BB/L011522/1). We thank Stephen Thorpe for writing a Python script that we used to compare the overlaps of the outputs of the terminator prediction algorithms.

One of the earliest industrial biotechnology processes used by man was the production of acetone by bacterial fermentation. The bacterium used was from the genus *Clostridium*, which are Gram-positive, spore forming, obligately anaerobic organisms¹. In addition to acetone, these processes produced butanol and ethanol, leading to the fermentation being given the name 'ABE' for 'acetone-butanol-ethanol'². ABE-producing clostridia typically show a biphasic growth pattern, producing acetic and butyric acids during the early stages of growth, and then undergoing a metabolic 'switch', reassimilating the produced acids and producing solvents³. Driven by the demands of industry, many different ABE-producing clostridial isolates were identified; those strains that were used in industry are primarily represented by four species^{4,5}, namely *Clostridium acetobutylicum*, *Clostridium beijerinckii*, *Clostridium saccharobutylicum*, and *Clostridium saccharoperbutylacetonicum*, although ABE fermentation has been observed in other clostridia⁶⁻⁸. Furthermore, not all butanol-producing clostridia produce the canonical trio of solvent products; for example, *Clostridium pasteurianum* produces butanol and 1,3-propanediol (instead of acetone) from glycerol⁹. While the ABE process had until recently fallen out of favour due to competition from the petrochemical industry, the necessity of identifying alternative fuels has renewed interest in the production of butanol as a potential biofuel candidate and sustainable commodity chemical¹⁰.

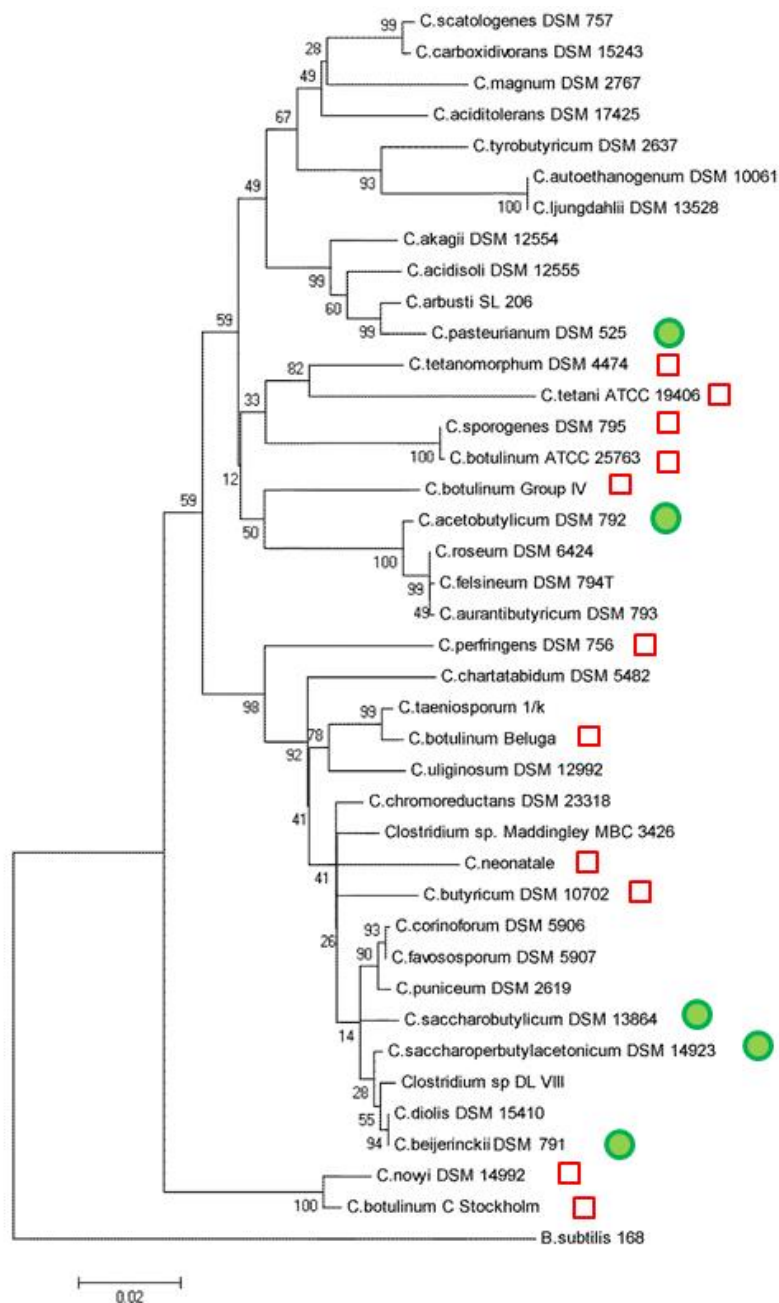


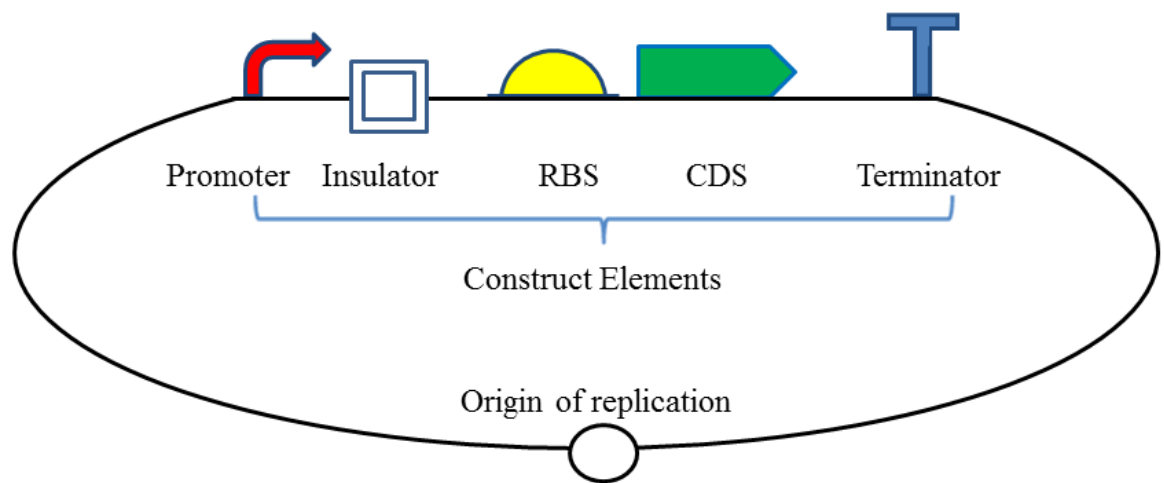
Figure 1. Phylogenetic tree of 16S sequences from selected organisms from the genus *Clostridium*. The tree was built using Maximum Likelihood method based on the Tamura-Nei model¹¹. The bootstrap consensus tree inferred from 1000 replicates¹² is taken to represent the evolutionary history of the taxa analyzed¹². Evolutionary analyses were conducted in MEGA5¹³. Green circles denote butanol-producing species used in industrial biotechnology (IB), red squares mark risk group 2 species¹⁴ (risk group 3 being highest risk).

A recent comparative genomics study of many industrial saccharolytic strains (those preferring sugar as a carbon source) revealed that the known solventogens fall within two sister clades: one exemplified by *C. acetobutylicum* and one by *C. beijerinckii*¹⁵ (Figure 1). Interestingly, another comparative genomics paper which included more genus-wide species but fewer industrial solventogens supported the split, with the genus' type species *Clostridium butyricum* being more closely related to *C. beijerinckii* and the pathogen *Clostridium tetani* clustering closer to *C. acetobutylicum*¹⁶. Altogether, these findings serve to re-iterate that complex traits within the Clostridia such as pathogenicity are paraphyletic (also see Fig. 1). On the other hand, solventogenesis (of ethanol, butanol or acetone) may be very widespread in the genus, but there has not yet been a definitive comparative study reporting the extent of its conservation to our knowledge, and species and strains certainly vary in their productivity¹⁷. Topics of engineering interest have included improved characteristics such as solvent production^{18,19}, sugar utilisation^{20,21}, growth on alternative feedstocks such as lignocellulose²²⁻²⁵ and the production of alternative products²⁶⁻²⁸.

A shortage of biological parts for the solventogenic Clostridia

To manipulate the solventogenic Clostridia using rational engineering, the development of reliable standard genetic parts for use in synthetic biology is essential. The assembly of these parts into synthetic gene cassettes and larger gene clusters means that they must work together in a consistent and predictable manner to become a useful tool for strain engineering. The basic parts of synthetic biology are the minimal sequence elements with biological function in gene expression (Figure 2), including promoters, ribosome binding sites, transcriptional terminators and other factors in the mRNA that affect stability. Also, synthetic biology tools rely on gene reporters to measure levels of gene expression, and these elements are often combined into plasmids which need their own origins of replication. Work in *Escherichia coli*, *Saccharomyces cerevisiae* and other model organisms has been transformed in the past decade by the rapid development, characterization and standardization of parts. However, work in these organisms benefits from a legacy of biological knowledge that is not necessarily available when working in non-model organisms, making the engineering of the latter even more

124 challenging.



125
126 **Figure 2.** Features of genetic constructs depicted using symbols from the SBOLv²⁹. Examples of promoters, ribosomal
127 binding sites, reporters, terminators and replicons are described in the main text.

128
129
130 It is also important to note that, of the existing body of molecular biology research in this
131 genus, the majority has been carried out in *C. acetobutylicum* ATCC 824. Work in this
132 species has benefited from a relatively early development of a reliable electroporation
133 protocol³⁰ and from the publication of the whole genome sequence in 2001, making it
134 the first published clostridial genome³¹. However, *C. acetobutylicum* is only one of many
135 industrially promising solventogenic clostridia. While the engineering of some of these
136 species has historically been hindered by a lack of transformation protocols, these are
137 now available for the transformation of the aforementioned five industrial species^{30,32–}
138 ³⁵. In several cases, development of transformation protocols has required the
139 circumvention of restriction systems which degrade incorrectly methylated DNA.
140 Indeed, electroporation of *C. acetobutylicum* only became viable with the expression of
141 the *B. subtilis* phage ϕ 3T I methylase in the *E. coli* cloning host; this methylates the
142 sequence GCNGC, which would otherwise be cleaved by the *C. acetobutylicum* Cac824I
143 type II restriction enzyme³⁰. Likewise, transformation of *C. pasteurianum* ATCC 6013
144 (DSM 525) requires the methylation of CGCG sequences, which has been accomplished
145 by the use of the M.FnuDII³³ or M.BepI³⁶ methylases. Other organisms have more
146 complex restriction systems; in *C. saccharobutylicum* NCP 262, which has two type I
147 restriction systems, expression of the methylation and specificity domains of these

systems on a plasmid in *E. coli* was sufficient to allow transformation by conjugation³⁴. Some developments have also been made in transforming non-type strains, which may have restriction patterns which differ from those of the type strains. For example, *C. pasteurianum* NRRL B-598 is part of the *C. beijerinckii* cluster¹⁵ but requires the use of a *dam*⁻/*dcm*⁻ strain of *E. coli* for successful transformation, suggesting that the type IV system of this strain is particularly important³⁷; conversely, *C. beijerinckii* NCIMB 8052 can be transformed with much greater efficiency even with DNA from a *dam*⁺/*dcm*⁺ host³⁸. While an analysis of such developments in the entire genus *Clostridium* would be beyond the scope of this review, recent publications by Pyne et al.³⁹ and Minton et al.⁴⁰ provide a comprehensive review of the development of *Clostridium* strains for genetic engineering. However, it is certain that the range of genetically tractable *Clostridium* species and strains will expand with future research.

The resurgent interest in solventogenic *Clostridium* species suggests that synthetic biology tools are needed and there are clear shortages of particular biological parts with characterized activity available for the engineering of these organisms. We believe this warrants a concerted effort to address the weaknesses in the toolbox. Furthermore, with the establishment of transformation protocols throughout the genus *Clostridium*, it is also important to consider the potential of adapting existing parts and tools for use in other, less well-developed solventogens. Here, we assess the currently available parts used in synthetic biology projects, starting first with reporters available to measure gene expression, then the promoters, terminators working at the level of transcription and then post-transcriptional factors including mRNA stability, ribosome binding sites, riboswitches and codon usage. Finally we review the replicons currently being used to assemble synthetic gene clusters and briefly provide an overview of methods to introduce these genetic elements onto the chromosome, which has been reviewed elsewhere⁴⁰. Overall, we aim to summarize and systematize the existing biological parts used in the industrial solventogenic *Clostridium* species in order to highlight areas where more research is required and accelerate progress in creating a full range of synthetic biology tools for manipulating these industrially important organisms.

Reporters of gene expression

Genetically-encoded reporter systems are the major *in vivo* gene expression measurement techniques available and are required to measure the activity of different parts being tested. The oxygen sensitivity of obligate anaerobes such as the clostridia limits the ease of use of many reporter systems, including some popular fluorescence-based and enzymatically-based reporters, and has led to the development of some novel reporters which are slowly being adopted.

Fluorescent Reporter Proteins

The use of fluorescent reporter proteins is now widespread in biology. Successful use requires the correct folding and maturation of the fluorescent protein to enable detection. The level of signal for the fluorescent protein must be sufficiently high to enable accurate detection as there is no signal amplification as seen in enzymatic reporters.

GFP-family fluorescent proteins

The highly engineered family of Green Fluorescent Proteins (GFPs) now includes variants with improved brightness and photostability and with a range of different colours (different excitation and emission maxima)⁴¹. The major limitation of the GFP-like proteins for anaerobes is the requirement of molecular oxygen for chromophore maturation and fluorescence⁴². This excludes the possibility of real-time gene expression monitoring; however, *in vivo* observations are still possible in some oxygen-tolerant clostridia such as *Clostridium perfringens*⁴³, where yellow fluorescent protein (YFP) was synthesized and fluorescence was developed after exposure of live cells to atmospheric oxygen⁴⁴.

Using a similar principle, it is possible to obtain a snapshot of the protein levels in whole cells by exposing fixed anaerobically grown bacteria to atmospheric O₂ in a process termed 'aerobic fluorescence recovery', enabled by the discovery that GFP-family proteins can undergo maturation in fixed cells. Two studies have demonstrated the technique in *C. difficile* using paraformaldehyde and glutaraldehyde to fix cells expressing codon-optimized variants of CFP (cyan)⁴⁵ and mCherry (red)⁴⁶. This approach works as the fixatives cross-link primary amines (especially lysine residues) which are absent from the GFP-family chromophores. Fixed cells offer an advantage to

the imaging of dying live cells as they more accurately represent normal protein localization (gene expression changes are also likely to occur in dying or metabolically stressed cells)⁴⁶. So far, the technique has not been used to systematically quantify gene expression but rather to label proteins and track their intracellular localization^{46,47}.

Flavin-binding Fluorescent proteins

The flavin-binding fluorescent proteins (FbFPs) are a class of alternative fluorescent reporters capable of maturation in anaerobic conditions. The FbFPs are small proteins (~11-15kDa) that have oxygen-independent fluorescent properties, using a flavin-mononucleotide (FMN) cofactor as the chromophore.

A pioneering study in 2007 reported the development of three anaerobic fluorescent reporters derived from the LOV (Light-, Oxygen- or Voltage-sensing) domains of bacterial proteins: BsFbFP derived from YtvA of *Bacillus subtilis*, PpFbFP from SB2 of *Pseudomonas putida* plus an *E. coli* codon optimized variant of BsFbFP called EcFbFP⁴². This domain was engineered to emit fluorescence by mutagenesis of a reactive cysteine in the FMN-binding pocket⁴².

These proteins are now commercially available from Evocatal GmbH under the trademark name 'evoglow'. Currently listed variants marketed for use in *Clostridium* species include Bs1 (monomeric BsFbFP), Bs2 (dimeric BsFbFP) and Pp1 (PpFbFP).

The reporters' functionality in *C. acetobutylicum* was demonstrated by Schulz in 2013; the highest fluorescence levels were observed for a *C. acetobutylicum* codon-optimized Pp1 (referred to in text as "Pp2"), followed by codon-optimized Bs2 (referred to as "Bs3"). Interestingly the same constructs gave inverse results in *E. coli*⁴⁸. While Evocatal GmbH offers C-Pp1 and C-Bs2 *Clostridium* codon-optimized reporters with publicly available nucleotide sequences, it is not clear whether the nucleotide sequences (reflective of the codon optimization approach) are the same as Pp2 and Bs3. A study in 2014 reported the placement of the evoglow Pp1 reporter downstream of the *cipP* promoter to monitor growth of *Clostridium cellulolyticum* on cellulose⁴⁹. The functionality of the evoglow Bs2 variant has also been demonstrated in *C. pasteurianum*⁵⁰ and in *Clostridium ljungdahlii*⁵¹.

Another FbFP example, developed from the *Arabidopsis thaliana* LOV2 domain of the blue-light receptor protein Phot2, is the improved LOV (iLOV) FbFP⁵². The iLOV FbFP has been further modified for enhanced properties, generating variants such as

photostable iLOV 2.1 (phiLOV 2.1) and others^{53,54}. While phiLOV 2.1 FbFP has improved photostability, it is still substantially lower than that of GFP. A recent study demonstrated the utility of a *C. difficile* codon optimized phiLOV 2.1 in three *Clostridium* species (expression was driven from an engineered strong *xyl* promoter)⁵⁵. Under the test conditions, the three species exhibited varying levels of fold-increase of fluorescence over their background autofluorescence level: 3.2-fold increase in *C. difficile* R20291 (not a *Clostridium sensu stricto* species), 5.6-fold in the pathogenic toxin-producer *Clostridium sordellii* ATCC 9714 and 4.5-fold in the solventogen *C. acetobutylicum* ATCC 824⁵⁵. While the FbFP reporters will be detectable when expression is driven from very strong promoters, further improvements in brightness and photostability will make them more generally useful for multiple applications.

One of the drawbacks of the FbFPs is that all of the available variants are of the same colour; additionally, unfolding caused by translocation has been reported to result in loss of fluorescence due to the loss of the bound flavin cofactor⁵⁴. These limitations prevent the use of FbFPs as the sole reporters for multi-output circuits or secretion. Engineering of FbFPs for different emission spectra has proven to be somewhat complicated; while a simulation study of a theoretical iLOV mutant predicted a red shift increase⁵⁶, experimental results showed that the mutation resulted in an overall brightness decrease and a blue shift of emission⁵⁷. Research is underway to develop more fluorescent derivatives of the LOV domains and to explore the potential for diversifying their colour spectrum^{56,57}. The other limitation of FbFPs is their relative dimness, which can decrease the sensitivity of a reporter system; depending on the measurement instrument used, a promoter's expression level might be below the detection limit. Photostability improvements (as in phiLOV) have resulted in a dimmer mutant⁵³; for applications where a longer exposure time is crucial, the dimmer, stable protein is preferable, but for an automated single measurement application such as flow cytometry, the brighter variant may be more suitable. Overall, all fluorescent protein reporters provide the ability to quantify cell-to-cell heterogeneity of gene expression (when used in flow cytometry or microscopy), which can be very useful, and while there are still some limitations with FbFPs, they have the potential to be a route to reliable *in vivo* real-time expression monitoring in *Clostridium*.

Enzymatic Reporter Proteins

Enzymatic reporters catalyse a (preferably) unique reaction either *in vivo* or *in vitro* and the specific activity is calculated to estimate protein levels and thus gene expression. Enzymatic reporters require the addition of substrates and cofactors to the assay reaction mixture as well as the production of cell lysate. This often means that enzymatic assays involve more preparation steps than fluorescent proteins but can have adjustable sensitivity by varying substrate levels. Also, *in vitro* lysate assays are, by definition, bulk population measurements.

Chloramphenicol acetyltransferase

The first reporter which has been extensively used in the *Clostridium* genus, both *in vivo* and, more quantitatively, *in vitro*, is the chloramphenicol acetyltransferase (CAT) reporter (encoded by *catP*). The system was first developed for use in *C. perfringens* (a medically relevant non-solventogenic bacterium)⁵⁸ and has since been used in *C. acetobutylicum*⁵⁹. Chloramphenicol acetyltransferase is an enzyme that catalyzes the covalent attachment of an acetyl group from acetyl-CoA to chloramphenicol⁶⁰, and is the basis of the chloramphenicol resistance marker found in many bacterial vectors. Transfer of the acetyl group exposes the thiol group of CoA, allowing the progress of the reaction to be observed by addition of 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, also known as Ellman's reagent); this compound reacts with the free thiol, releasing 5-thio-2-nitrobenzoate, which can be detected by measurement of absorbance at 412 nm^{60, 61}. The assay relies on a continuous spectrophotometric rate determination to calculate specific activity. Potential drawbacks include endogenous activity in chloramphenicol resistant strains (which may be resolved by disruption of the resistance gene) as well as high levels of endogenous non-specific coenzyme A transferases⁶² (knockouts of which would be more laborious and would likely have growth and phenotypic effects) and the cost of the substrate acetyl-CoA.

Carbohydrate hydrolases: β -glucuronidase, β -galactosidase, amylase, endoglucanase

The CAT assay's drawbacks led to the adaptation of a classical *E. coli* reporter in *Clostridium*: the β -galactosidase enzyme, encoded by *lacZ* (the gene was derived from *Thermoanaerobacter thermosulfurigenes*)⁶². To quantify enzyme activity spectrophotometrically, ortho-nitrophenyl- β -galactoside (ONPG) is used in an *in vitro* assay and an increase in absorbance at 420 nm due to the release of ortho-nitrophenol is

measured. Similarly, the GUS reporter system, which utilizes β -glucuronidase (*E. coli gusA*), has been used in *C. acetobutylicum* in a fluorimetric assay with a cell lysate⁶³. The fluorimetric assay measures the release of 4-methylumbelliferone (4-MU) after cleavage of 4-MU- β -galactoside or 4-MU- β -glucuronide (by a β -galactosidase or β -glucuronidase, respectively), 4-MU emits light at 460 nm when excited by 365 nm light⁶⁴. Both the β -galactosidase and β -glucuronidase reporter systems benefit from the commercially available range of fluorometric, spectrophotometric and histochemical substrates and kits, making them an improvement over the CAT assay, although they are still not inexpensive.

The endogenous *amyP* gene (encoding an amylase expressed during solventogenesis) has been used in *C. acetobutylicum*⁶⁵ as a reporter to study the phenomenon of strain degeneration (loss of solventogenesis) which is often caused by loss of the pSOL1 megaplasmid on which *amyP* is located. A codon-optimized amylase (*AmyE^{opt}*) has been used successfully as a secreted reporter in *C. difficile* by addition of a zinc metalloprotease PPEP-1 signal sequence⁶⁶. It is noteworthy that the strain used in the above study was not capable of degrading starch under laboratory conditions; use of amylase as a reporter in amylolytic strains (such as *C. acetobutylicum* ATCC 824) may require knockout of endogenous amylases to increase signal to background ratio.

In *C. beijerinckii*, a secreted endoglucanase (*eglA*) was cloned from *C. saccharobutylicum* NCP 262 and used as a reporter for gene expression⁶⁷; assays used were agar plate and cell lysate assays which measured substrate (carboxymethylcellulose, CMC) clearance on plates and product (p-nitrophenol from cleavage of p-nitrophenyl cellobioside) accumulation in lysate, respectively.

Lipase and Alkaline phosphatase

The lipase encoded by *tliA*, from *Pseudomonas fluorescens* SIK W1, has also been used as a reporter in *C. beijerinckii* NCIMB 8052⁶⁸. Lipases are enzymes which hydrolyse the ester bonds found in long-chain acylglycerols, releasing fatty acids⁶⁹. Activity can thus be assessed by an enzymatic assay measuring the cleavage of p-nitrophenyl-esters⁶⁹ such as p-nitrophenyl decanoate⁶⁸. However, this reporter was only useable in *C. beijerinckii* as this species has a very low endogenous lipase activity. As such, it may not be suitable in other species, such as *C. acetobutylicum*, which has been observed to show inducible lipase activity⁷⁰.

A colorimetric alkaline phosphatase assay was developed and used in *C. difficile* in 2015 based on the *phoZ* gene product from *Enterococcus faecalis*⁷¹. Activity of this gene can be determined by a colorimetric assay with *p*-nitrophenyl phosphate as the substrate. Use in solventogenic clostridia could be limited by native phosphatase activity. In order to examine the suitability of *C. difficile* as a host for this reporter, BLAST analysis was used to screen for *phoZ* homologues, and activity towards 5-bromo-4-chloro-3-indolyl phosphate was tested⁷¹. While the four main industrial strains do not have a *phoZ* homologue, homologues can be found in other species such as *C. pasteurianum* (CLPA_RS02340, with 29% identity to *phoZ*).

Luciferase

The luciferase (*lucB*) reporter was also used in *C. acetobutylicum* successfully, allowing luminescence detection⁷². However, it also requires oxygen, ATP and luciferin. Notably, the cells used in the luciferase assay were neither lysed nor fixed, but were live cells that were exposed to atmospheric conditions, washed, and kept on ice. This treatment could conceivably introduce changes in gene expression levels prior to measurement. This assay has the lowest background signal level but the requirement for live cell exposure to oxygen may introduce variability. A codon-optimized luciferase (*sLuc*^{opt}) was also successfully secreted in *C. difficile* using the aforementioned zinc metalloprotease signal peptide⁶⁶.

While there are many reporter choices available to clostridial researchers, we would argue that the multiplicity of reporters used has not helped ease the comparison of data obtained by different laboratories. Altogether, a single reporter has not been established as a community standard; given the drawbacks of each particular system, it is difficult to identify one standout reporter, although our hope would be that improved fluorescence reporters with increased brightness and photostability would be the most useful and enable single cell studies in live cells. Currently we would advocate a choice of more than one reporter, enzymatic and fluorescent, and their systematic and comparative use within the genus. Such an approach could provide insight into the sources of gene expression variability between strains and provide a starting point for future engineering. To achieve this seamlessly, it is advisable to explore translational fusions between reporters; examples from *E. coli* include a FRET pair YFP-FbFP fusion⁷³ and

Gemini (lacZ α -GFP) fusion⁷⁴. If successfully applied to the *Clostridium* species, such bi-functional reporters have the potential to become a single standard reporter. Additionally, α -peptide complementation strategies using the *E. coli lacZ* gene in *B. subtilis*⁷⁵ and *Mycobacterium smegmatis*⁷⁶ have been reported but a similar approach in Clostridia has not been attempted. Interestingly, the *T. thermosulfurigenes*-derived LacZ protein appears to lack the α -complementation region⁷⁷ and is a dimer⁷⁸ (rather than a tetramer like *E. coli* LacZ).

Transcription

Transcription is the first stage of gene expression and the main stage for regulation of gene expression. Most existing biological parts in *Clostridium* fall within the transcription category and are mostly promoters (summarized in Table 1). In contrast, the termination of transcription has been studied relatively little and we present some analysis on the potential to study and improve the parts available for reliable termination.

Transcription initiation

Promoters are DNA elements that are capable of driving transcription by RNA polymerase (RNAP) of downstream regions. Core promoter architecture in bacteria includes the -10 region or TATAAT box (Pribnow sequence), the -35 region and a spacer (with an optimal length of 17bp in *E. coli*) between the two conserved regions⁷⁹. This motif is recognized by the housekeeping sigma factor that provides sequence specificity to RNAP (called RpoD in *E. coli* and SigA in *B. subtilis*, *C. acetobutylicum* and other solventogens)^{80,81}. Both the consensus sequence and sigma factors bear very significant similarities between Firmicutes and *Escherichia*.

In addition, some *E. coli* promoters contain AT-rich UP elements (upstream of -35 region) that are responsible for recognition by the carboxy-terminal domain of the RNAP α -subunit⁸², an additional *E. coli* promoter feature is the 'extended -10' region⁸³ (upstream of the -10 region and within the spacer). Interestingly, near-consensus promoters (including ones with UP elements) have been found to be significantly more common in Firmicutes (including Clostridia) than in other bacteria⁸⁴, a feature that was not explicable through higher AT-content alone. In Clostridia additional sequences with resemblances to UP elements (termed phased A-tracts) have been described⁸⁵ and a conserved extended -10 region that differs from its *E. coli* analogue has also been proposed⁸⁶. The majority of promoters used in the genetic engineering of clostridia have

been identified from the transcriptional units of important metabolic genes and most have been shown to be able to drive strong constitutive expression of a gene of interest⁸⁷. The most commonly used constitutive promoter in *C. acetobutylicum* (Table 1) is that of the native thiolase gene (*thlA*)^{88,89}, which has been used for the expression of genes involved in production of solvents such as butanol⁹⁰ and isopropanol²⁶, the transcriptional regulator gene *tetR*⁹¹ (used a minimal promoter variant- *miniPthl*), and, in a modified form, for the expression of cellulosomal scaffoldins⁹² and glycoside hydrolases^{25,93}. The *C. acetobutylicum* thiolase enzyme catalyzes the condensation of acetoacetyl-CoA from acetyl-CoA which are central metabolic intermediates⁸⁸. The *C. acetobutylicum thlA* promoter (hereinafter referred to as *thlA*) is a sigA-dependent promoter (as evidenced by its near-consensus -35 and -10 regions⁸⁹); however, in its normally chromosomal context it is also the subject of regulation by the redox-sensing transcriptional repressor Rex⁹⁴. Rex inactivation was found to increase native *thlA* activity in *C. acetobutylicum* about 12-fold⁹⁵. The Rex-binding site has been omitted from the core promoter in commonly used synthetic constructs⁹⁶; yet this binding site is conserved in the promoters of orthologous genes from other solventogens (according to the RegPrecise database⁹⁷ and our promoter region alignments), suggesting that there are additional levels of control for this promoter that could be easily added by addition of the Rex sites if desirable. Indeed, some studies using similar promoters such as *thlA* and *ptb* have reported conflicting strength/activity findings, which may well be due to the cloning of regions of varying length (that likely include regulator binding sites such as Rex) and choosing to include the native RBSs or not (such as the ones reported between promoters in the Schulz 2013⁴⁸ and Girbal 2003⁶³ studies).

Thiolase promoters have also been used for expression in other clostridia such as *C. beijerinckii*⁹⁸ and *C. pasteurianum*⁹⁹. However, gene expression data (RNAseq from *C. beijerinckii*) indicates that there are other genes that have higher expression levels than the ones currently used as promoter sources; many of those genes encoded are hypothetical proteins¹⁰⁰. Recently, Yang *et al.*, 2017 constructed a sequence logo of 18 promoters previously identified in *C. acetobutylicum* which revealed a strongly conserved -10 region. Using degenerate oligos to mutagenize the core *thlA* promoter elements, the authors were able to generate a promoter library of variable strengths,

including mutations that increased the strength of the promoter, suggesting that *thlA* can be improved further using synthetic biology approaches.

The use of a constitutive promoter may not always be desirable; it may be preferable to use an inducible promoter, allowing controlled expression of a gene of interest. So far, the only naturally inducible promoters exemplified in a solventogenic clostridium are the *Staphylococcus xylosus xylA* promoter, which is repressed by XylR in the absence of D-xylose⁶³, and the *C. perfringens bgaL* promoter, which is repressed by BglR in the absence of lactose¹⁰¹. However, constitutive promoters can be made inducible by addition of operator sites (for transcriptional repressors) or by the addition of binding sites for activators, and several such promoters have been developed for use in solventogenic clostridia. LacI-repressible versions of thiolase and ferredoxin promoters have allowed the construction of clostridial expression constructs for genes where expression in *E. coli* would be toxic^{24,25,93}. In *C. acetobutylicum*, the *fac* promoter (*Clostridium pasteurianum* ferredoxin promoter with *lac* operator) is able to function as an IPTG-inducible system in the presence of LacI, allowing approximately 10-fold induction⁵⁹. A TetR-repressed, anhydrotetracycline-inducible promoter has also been used in *C. acetobutylicum*, generated by the fusion of the chloramphenicol acetyltransferase promoter *pcm* with the tetracycline operator *tetO*⁹¹. This promoter could achieve up to 313-fold induction, although high levels of anhydrotetracycline were inhibitory to growth. However, operators are known to influence the basal activity of the promoter¹⁰². In fact, a recent study in *E. coli* found the core RpoD promoter too sensitive to sequence context and operator insertions to be a suitable target for forward engineering efforts and turned to extracellular sigma factors (ECFs) with T7 RNAP whose promoter core sequences they found to be more insensitive to operator addition¹⁰³. An alternative strategy - CRISPR-mediated repression of transcription - has been demonstrated in several solventogenic species, namely *C. acetobutylicum*¹⁰⁴, *C. beijerinckii*⁹⁸ and *C. pasteurianum*⁵⁰,

Orthogonal expression systems, i.e. promoters that require other non-native elements for activity (commonly an alternative sigma factor or phage polymerase), first found widespread use in the cloning of genes with toxic products. This approach allows the total repression of genes until they are introduced into the organism of interest. A

477 commonly used example in *E. coli* is the T7 promoter, which requires the phage T7
478 polymerase for activity¹⁰⁵. An example from Clostridia is expression driven from the *C.*
479 *difficile tcdB* promoter that is dependent on the native sigma factor TcdR. The *tcdB*
480 promoter is highly active in *C. acetobutylicum*, but only when *tcdR* has been integrated
481 into the genome⁵⁹. This enabled the high-level expression of a *mariner* transposon in *C.*
482 *acetobutylicum* without negative effects on the *E. coli* cloning host. There is a strong case
483 for designing these orthologous-type systems as simply using a clostridial
484 housekeeping promoter (such as the *thlA* promoter) with a clostridial ribosome binding
485 site would not provide orthogonality in commonly used cloning hosts such as *E. coli* or *B.*
486 *subtilis* due to the degree of conservation between the transcription and translation
487 initiation systems amongst these organisms; on the contrary, as a rule strong clostridial
488 promoters and RBSs (see translation initiation section) often retain their strength in *E.*
489 *coli* whereas the opposite is observed more rarely, this is likely due to the on average
490 higher similarity to the Bacteria-wide consensus translation and transcription initiation
491 signals in *Clostridium*^{86,106}. Interestingly, AT-rich DNA can be toxic to *E. coli* (clostridial
492 genomes are very AT-rich) due to transcriptional activity from spurious intragenic
493 promoters and RNAP titration¹⁰⁷. This is exacerbated by the fact that the *E. coli* extended
494 -10 region, which is sufficient alone for transcription initiation in *E. coli*, could also be
495 present by chance in *Clostridium*-derived sequences, increasing the likelihood of
496 spurious and unpredictable transcription during cloning in *E. coli*.

Table 1. Promoters used in engineering of *C. acetobutylicum*

Promoter	Comments
Constitutive	
<i>thlA</i> (<i>C. acetobutylicum</i>)	Widely used for constitutive gene expression in <i>C. acetobutylicum</i> for example expression on <i>C. beijerinckii ald</i> , <i>C. ljungdahlii bdh</i> ⁹⁰ , <i>C. acetobutylicum adc</i> , <i>ctfA</i> , <i>ctfB</i> , ²⁶ & <i>tetR</i> (<i>miniPthl</i> variant) ⁹¹ . Activity analyzed using Gusa ⁶³ and several FbFP ⁴⁸ reporters
<i>thl</i> (<i>C. perfringens</i>)	Used for expression of <i>srtA</i> genes from <i>C. acetobutylicum</i> , <i>L. monocytogenes</i> , and <i>B. cereus</i> ²⁵ .
<i>fdx</i> (<i>C. sporogenes</i>)	Activity analyzed using CatP reporter ⁵⁹ . Used in the CloStron system for expression of the Ll.LtrB intron ¹⁰⁸ . Used for expression of <i>spoA</i> integrated into the chromosome for complementation of a <i>spoA</i> mutant ¹⁰⁹ .
Controlled	
<i>thl</i> (<i>thl</i> with double <i>lac</i> operator)	<i>lac</i> -repressed version of <i>thlA</i> . Used to express a miniscaffoldin <i>cipc1</i> ⁹² & weakened version for expression of mannanase <i>man5K</i> ⁹³ .
<i>thlOid</i> (<i>thl</i> with single <i>lac</i> operator)	A <i>lac</i> -repressible version, used to drive chromosomal expression of the <i>C. cellulolyticum</i> glycoside hydrolase <i>xyn10A</i> ²⁵ .
<i>fac</i> (single <i>lac</i> operator, derived from <i>C. pasteurianum</i>)	Activity analyzed using CatP reporter ^{59,110} and Pp2 FbFP ⁴⁸ . Formerly used in the CloStron system for expression of the Ll.LtrB intron ¹¹⁰ . Used for expression of <i>codA</i> in the <i>C. acetobutylicum</i> knockout vector pMTL-SC7515 ¹⁰⁹ .
<i>fdxOid</i> (<i>fdx</i> with single <i>lac</i> operator)	Used for chromosomal expression of <i>C. cellulolyticum</i> glycoside hydrolase <i>cel9G</i> ²⁵ . Has RBS from <i>C. acetobutylicum thlA</i> promoter.
<i>facOid</i> (<i>fac</i> with single <i>lac</i> operator)	Used for chromosomal expression of <i>C. cellulolyticum</i> glycoside hydrolase <i>cel48F</i> ²⁵ . Has RBS from <i>C. acetobutylicum thlA</i> promoter.
<i>tcdB</i> (<i>C. difficile</i>)	Developed for an orthogonal expression system, requires the exogenous sigma factor TcdR for function ⁵⁹ .
<i>xylA</i> (<i>S. xylosus</i>)	Repressed by XylR and inducible by D-xylose; assessed with β -glucuronidase reporter giving 17-fold induction with D-xylose ⁶³ .
<i>pcm</i> (<i>tetO1</i> -containing variants)	Repressed by TetR; inducible in the presence of anhydrotetracycline ⁹¹ .
<i>bgaL</i> (<i>C. perfringens</i>)	Repressed by BgaR and inducible by lactose ¹⁰¹ .
<i>xyl/tetO</i> (<i>B. subtilis</i>)	Repressed by TetR ¹¹¹ ; activity in <i>C. acetobutylicum</i> shown by phiLOV 2.1 Opt FbFP expression ⁵⁵
Fermentation phase-specific	
<i>adc</i> (<i>C. acetobutylicum</i>)	Activity assessed using β -glucuronidase ⁶³ , β -galactosidase ⁷² and Pp2 FbFP ⁴⁸ . Primarily active after onset of solventogenesis.
<i>ptb</i> (<i>C. acetobutylicum</i>)	Activity assessed using β -glucuronidase ⁶³ , luciferase ⁷² and Pp2 FbFP ⁴⁸ . Active during acidogenesis (not solventogenesis)
<i>ptb</i> (<i>C. beijerinckii</i>)	Activity assessed through expression of <i>lacI</i> ⁵⁹ .
<i>sol</i> (<i>C. acetobutylicum</i>)	Activity assessed using β -galactosidase & luciferase reporters ⁷² . Weak expression, primarily active during late exponential phase.
<i>bdhA</i> (<i>C. acetobutylicum</i>)	Analyzed using β -galactosidase reporter ⁷² . Comparatively weak expression, primarily active in early exponential phase.
<i>bdhB</i> (<i>C. acetobutylicum</i>)	Analyzed using β -galactosidase reporter ⁷² . Primarily active until onset of solventogenesis.
<i>hydA</i> (<i>C. acetobutylicum</i>)	Activity analyzed using β -glucuronidase ⁶³ and Pp2 FbFP ⁴⁸ . High activity during acidogenesis, decreases to low after phase shift
<i>adhe2</i> (<i>C. acetobutylicum</i>)	Activity analyzed using Pp2 FbFP ⁴⁸ . Strong expression during solventogenesis ^{48,112}

Transcription termination

Bacteria have two distinct mechanisms that function in transcription termination. Both types of terminators are usually located in the 3' end of transcriptional units. Rho-dependent terminators rely on the Rho protein that recognizes a target sequence and causes RNA polymerase to fall off of the template DNA¹¹³. The specific DNA recognition sites, *rut* sites, have been used before in genetic circuits¹¹⁴, but not extensively, perhaps due to the relatively poor mechanistic understanding of the process¹¹⁵. The second mechanism, which also has been reported to be more widespread¹¹⁶, is referred to as Rho-independent or intrinsic termination and relies on the folding of a short GC rich hairpin followed by a poly-U transcribed sequence. The hairpin folding and subsequent transcription of the U-tract causes the polymerase to pause and leads to transcript and RNAP dissociation¹¹⁷. This is also facilitated by the fact that ribo-uracil-deoxyribo-adenine is the weakest nucleotide base pair¹¹⁸. Intrinsic terminators are often found downstream of operons; however, they are also involved in transcription attenuation when present within coding regions or downstream of promoters¹¹⁹.

While a number of terminators have been used in the construction of clostridial expression constructs, often derived from clostridial genes such as *adc*^{63,120}, *fdx*¹⁰⁹, CD0164¹⁰⁹ and *thlA*^{90,121}, there have been few published analyses of terminator strengths in clostridia. The *C. pasteurianum* *fdx* terminator has been shown to be highly effective at preventing read-through inhibition of the replicon from the *fac* promoter in a clostridial vector⁵⁹, and screening of a selection of terminators in *C. acetobutylicum*²⁵ showed that the *E. coli* *rrnB* terminator T1 loop was able to function as an efficient terminator, reducing expression of a downstream gene. However, these analyses only took into consideration the effect of the terminator on a downstream target under the influence of a single promoter. As previously mentioned, the introduction of a terminator may have an effect on expression of an upstream gene by influencing mRNA stability, which is not a desirable feature of a standard part¹²². However, this effect can be quantified in an appropriately designed assay¹²³. Furthermore, the efficiency of termination may increase or decrease depending on promoter activity; this has been recently exploited for the development of a genetic band-pass filter in *E. coli*¹²⁴. Terminator strength has also been shown to be influenced by the hairpin's proximity to a stop codon or when present within a coding region¹²⁵. A recent study demonstrated

the gradual increase in termination efficiency as distance between the stop codon and hairpin increased¹²⁶. The ribosome was found to repress transcriptional termination when the stop codon and hairpin were in close proximity – termination repression was relieved when insulator sequences of approximately 30 bp were used¹²⁶. Expanding the range of characterized clostridial terminators is clearly necessary for the development of a clostridial synthetic biology toolbox.

One approach that can be undertaken is the use of algorithms to extract putative terminator sequences from genomes, also known as ‘part mining’. Several such bioinformatics tools exist; they rely on seed sequences, secondary RNA structure features or both^{116,127–129}. These bioinformatics tools were developed to aid genome annotations by identifying the 3’ ends of operons and may not then detect all the features required by a functional terminator. In order to evaluate the relative difference in sequences retrieved by different programs, we compared the output of three searches performed on the *C. acetobutylicum* ATCC 824 genome (Figure 3A). The three search algorithms that were chosen are as follows: TransTermHP (TTHP)¹²⁸, RNIE¹¹⁶ and WebGeSTer (WG)¹³⁰. TTHP is a widely used tool, while RNIE’s authors state that their tool eliminates false positives in comparison to TTHP¹¹⁶. Finally, WebGeSTer was selected because the authors classify the results into different types of intrinsic terminators based on overall secondary structure. They discovered that canonical U-tract containing intrinsic terminators (termed ‘L-shaped’) form the majority of structures found within Firmicutes, whereas they are a minority in the *E. coli* model where they were first identified¹²⁹.

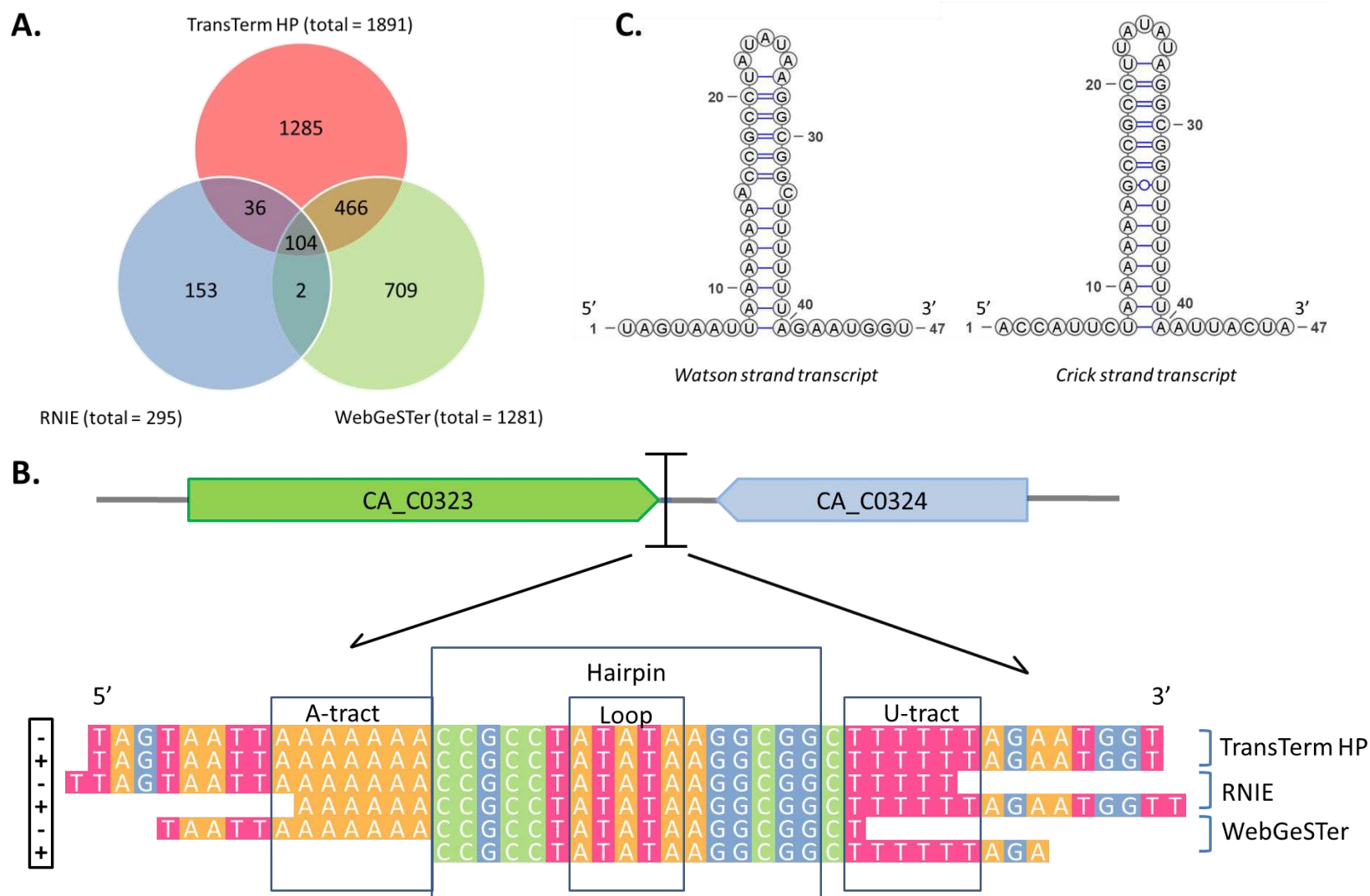


Figure 3. Application of terminator prediction software to *C. acetobutylicum*. **A.** Outputs from predicted terminator sets derived from the *C. acetobutylicum* genome by application of TransTermHP (TTHP), RNIE and WedGeSTer were compared. **B.** An alignment of predictions of a putative bi-directional terminator (marked by T sign) by all three algorithms (note all three produce separate predictions for plus (+) and minus (-)). **C.** RNA structures predicted by RNAfold^{215,216} for the TransTermHP sequence, drawn with VARNA²¹⁷; it is worth noting that the base-pairing between the A-tract and U-tract may not form until after termination²¹⁸. Terminator lengths: WG: maximum length (max)-48bp, minimum (min) - 28bp, average (avg) - 35.8bp; RNIE: max-45bp, min- 34bp, avg-41.7bp, TTHP: max-89bp, min-41bp, avg-57.2bp.

In our analysis, each algorithm was used to produce sets of putative terminator predictions, which were examined to identify overlapping predictions and are presented in Fig. 3A. Using the default settings of the different algorithms, the total number of predictions per program varied considerably, with TTHP and WG each predicting over 1000 terminators while RNIE predicted fewer than 300 (Fig. 3A). Despite this much smaller set, only about half of the predictions were shared with either TTHP and WG sets or both, producing a small set of 104 terminators shared by all three methods, an example of which is illustrated in Fig. 3B, an intergenic region separating two converging coding sequences, CA_C0323 (a sensory transduction histidine kinase) and CA_C0324 (a tetratricopeptide-repeats-containing protein). This also illustrates a class of terminators with particular utility in synthetic biology, namely bi-directional terminators, which are capable of terminating transcription from both the plus and minus strands. Our analysis finds that 37 bi-directional terminators are supported by all three algorithms, supporting the concepts that there are plentiful targets in clostridial genomes for 'part mining' and future experimental characterisation of these likely strong terminators. An interesting feature of the bi-directional terminators is their hypersymmetry¹³¹ (can be seen in Fig. 3B). This leads to the possibility to form an extended hairpin between the so-called A-tract (can act as U-tract in the reverse transcriptional orientation) and the U-tract (as shown in Fig. 3C). However, as pausing and termination occur at the U-tract¹¹⁷, the formation of an extended hairpin may not be relevant and does not appear to contribute to strength in all terminators¹³². Also, note that in this example the terminator on the minus strand (also referred to as Crick strand) is predicted to have a more stable hairpin (rG and rU form a basepair while rA and rC do not), whereas terminators from both strands are capable of forming an extended hairpin if fully transcribed as a result of base pairing between the A- and U-tracts. Combined with having a longer U-tract (Fig. 3C), the Crick strand terminator is possibly the stronger one, interestingly WebGeSTer predicts a smaller terminator downstream of the one depicted in Fig. 3A (not shown). A previous analysis of the *C. acetobutylicum* genome for terminators using TransTerm¹²⁷, a predecessor of TTHP, revealed the presence of 66 putative bi-directional terminators¹³³, supporting the rationale of combining outputs from multiple algorithms to create a small part list for experimental characterisation.

Translation

Translation initiation

Bacterial ribosome-binding sites (RBS) are short sequences located in the 5' untranslated region of messenger RNA (mRNA) transcripts, consisting of a Shine-Dalgarno sequence (SD), polynucleotide spacer, and a translation initiation codon¹³⁴. Commonly, native promoter-RBS combinations have been used (such as in the pMTL80000 vectors). An alternative strategy is to use a native RBS (such as that of the *C. acetobutylicum thlA* gene) fused to a new promoter, as in a study from 2016²⁵ that generated several new hybrid promoters. Others have experimented with the length of the spacer^{135,136}. Yet, in contrast to model organisms (*E. coli* and even *B. subtilis*), there are few published comparisons of modified RBSs for use in solventogenic clostridia. Hence the generation and screening of synthetic RBS libraries could be a promising route for optimising the expression of synthetic gene constructs in clostridia.

SD sequences provide sequence complementarity for the 3' terminus of the 16S rRNA (known as the anti-Shine-Dalgarno or aSD¹³⁷) which acts as a guide for the ribosome complex enabling mRNA recognition by the translation machinery and translation initiation. While most of the knowledge on bacterial translation initiation comes from *E. coli* work, early studies indicated that the Firmicute *B. subtilis* requires a longer complementary region between the SD and the 16S rRNA to achieve comparable expression levels^{138,139}. An early study on translation initiation revealed that Firmicutes have, on average, a higher complementarity of the predicted SD region to the 16S rRNA 3' terminus than *E. coli* does¹⁰⁶. Recently, a systematic analysis of SD-aSD pairings in *B. subtilis* and in *E. coli* confirmed this trend¹⁴⁰. Replicating these studies in solventogens would provide useful information.

The spacer is the mRNA region between the SD and the start codon. The aligned spacer (which is the distance between the start codon and the 5' end of the aSD, base-paired to a SD^{137,141,142} as revealed by sequence alignment) is of particular importance for translation initiation¹³⁷. Defining the length of the aligned spacer precisely is difficult, as interpretations of the extent of the aSD region (starting from the 3' 16S rRNA terminus) vary. The 3' end of the 16S rRNA gene (aSDs) of *B. subtilis* is identical to that of *C. perfringens*¹⁴³ (as well as identical to those of solventogenic clostridia, based on our sequence analysis of published genomes) but to our knowledge there is no reported

experimental validation of the clostridial mature 16S rRNA 3' ends *in vivo*. Spacers in different species may have different optimal lengths; for example, spacers of *Pyrococcus abyssi* are, on average, roughly 3 nucleotides longer than those of *E. coli*¹⁴⁴, whereas the spacers in *Bifidobacterium longum*¹⁴⁵ would be considered shorter. According to our definition of the putative clostridial aSD (5' GAUACCUCCUUUCU 3'), in *C. acetobutylicum*, the native RBS of the *thlA* promoter has an 'aligned spacer' of 4 bases.

The effect of altering the length of the *thlA* spacer was recently investigated in two studies conducted in *C. acetobutylicum*^{135,136}. Interestingly Yang *et al.*, 2016 showed that a lengthened *thlA* RBS spacer with an extra 6 nucleotides (to a total of 14 bases- aligned spacer of 10bp), encoding a Sall restriction site, did not significantly alter reporter expression in comparison to the WT *thlA* RBS. Shortening the spacer below the WT length resulted in a decrease in expression, while further increases over an aligned spacer length of 10 bases (by the addition of a XbaI site) removed almost all of the expression¹³⁵.

Yang *et al.* (2016) demonstrated the potential benefits of utilising modified spacers in *C. acetobutylicum*¹³⁵. Overexpression of the biotin synthesis genes *bioY*, *bioD*, *bioA*, and *bioB* was observed to provide an improvement in growth phenotype and solvent production. The *thlA* promoter was used to drive expression of *bioY*, *bioD* and *bioA*; replacement of the *thlA* RBS spacer with a shortened, less 'effective' variant resulted in a further improvement in growth characteristics. These results demonstrate not only the ability of modified RBS to optimize expression of synthetic pathways, but also the necessity of considering the effects of any alterations to the spacer, e.g. the introduction of restriction sites, when generating synthetic constructs. In 2017, Yang *et al.* generated a library of spacers (a BamHI site preceded the start codon in all cases) by starting with an aligned spacer of 2 bases and progressively increasing its length by 2 bases. After testing 11 variants the authors found that an aligned spacer of 4 bases gave the strongest levels of expression followed by 8, 6 and 10 bases in that order. It is conceivable that not only the length but also the sequence of the spacer might influence translation initiation, for example the introduction of a secondary SD within the primary SD's spacer region.

Computational tools to design RBSs exist, such as the RBS calculator, which use biophysical models of RBS recognition and translation initiation, including RNA base-pairing between the aSD and SD, spacer length and messenger secondary

structure^{146,147}. There are reports that *de novo* design produces more accurate results than translation initiation rate prediction of natural sequences for Gram positives¹⁴⁸.

Codon usage during translation

While codon usage is not formally a 'biological part', it is an important feature of coding sequences, the differential frequency of synonymous codons amongst genomes, referred to as codon usage bias^{149,150}, has been shown to strongly influence heterologous protein expression levels^{151,152}.

High genomic AT-content is characteristic of the Firmicutes and is reflected in the nucleotide composition of coding sequences. The *Clostridium* species' codon usage differs from that of other Firmicutes as well as the Proteobacterium *E. coli*¹⁵³ and there are also bioinformatically observable variations within the *Clostridium* genus itself¹⁵⁴ but the significance of the latter in influencing gene expression has not been experimentally verified to our knowledge. Genetically encoded reporters have been used heterologously within the Firmicutes phylum without codon-optimization, for example *Staphylococci* have been sources of reporters and antibiotic resistance genes for Clostridia. On several occasions researchers have successfully used native reporter genes from *E. coli* (*gusA*)⁶³, *T. thermosulfurigenes* (*lacZ*)⁷² and the firefly *Photinus pyralis* (*lucB*)⁷² (we describe these reporters in more detail in the Enzymatic Reporter Protein section). Codon-optimized genes for *C. difficile* have been used in *C. acetobutylicum*⁵⁵, as well as bespoke *C. acetobutylicum* codon-optimized ones^{24,48}. Researchers have also codon-optimized several GFP-like proteins for use in Firmicutes¹⁵⁵ (with *B. subtilis* in mind) but these have not been used in solventogens to our knowledge.

Codon optimization is not a trivial problem and codon optimization strategies vary considerably. The codon-adaptation index (CAI)¹⁵⁰ has been the historical measure of codon usage bias in an organism while there are others such as the codon bias index and the effective number of codons¹⁵⁶. Interestingly, simply improving a heterologous gene's CAI (making it more like a native gene or a highly expressed native gene) has not been found to correlate with expression levels^{157,158}. Efforts have been made to improve the indices describing codon usage bias and translational efficiency (which codon usage bias is thought to reflect) by studying endogenous gene expression^{159,160}. In addition, condition-specific usage tables have also been reported¹⁶¹. In an alternative strategy, Welch and colleagues developed genetic algorithms to select partial least squares

regression models which revealed that codons predominantly read by tRNAs that are most highly charged during amino acid starvation were good predictors of expression levels. Based on these results the researchers developed proprietary codon optimization algorithms to maximize protein expression^{158,162} which allowed them to predict expression levels in *E. coli*.

While it is clear that there is room for improvement in the heterologous protein expression strategies used in *Clostridium* solventogens, codon optimization strategies themselves are still being developed and the underlying principles are not yet fully understood; progress and existing approaches in the field have been reviewed elsewhere¹⁶³. A good starting point is for researchers to report the details of the codon optimization strategy undertaken when publishing work containing codon-optimized genes. This way data from heterologous protein expression in solventogens can be compared more reliably.

Post-transcriptional control of gene expression

Tuning gene expression levels in *Clostridium* species has been achieved using control at the RNA level – by either influencing translation or changing RNA degradation rates.

Controlling mRNA stability

Another potential avenue for optimisation of expression levels is the adjustment of mRNA stability. Altering the stability of an mRNA transcript influences the number of transcripts in the cell, thereby affecting the overall rate of translation. In bacteria, a number of factors are associated with mRNA stability, such as secondary structures, RNase recognition sites and polyadenylation, amongst others¹²². The presence of secondary structures at the 5' end of the mRNA has been observed to provide an increase in mRNA stability¹²². In *E. coli*, these structures prevent the binding of RNase E, an endonuclease which binds single-stranded RNA at the 5' end and then scans for cleavage sites. This property has been exploited in *E. coli* by the generation of libraries of synthetic hairpins for introduction into the 5' untranslated region (UTR)¹⁶⁴. While *C. acetobutylicum* has an RNase E homolog, RNase E/G, it is not certain if this behaves in the same way as *E. coli* RNase E, due to having a different domain organisation¹⁶⁵. Nevertheless, *C. acetobutylicum* also has a homolog of RNase Y¹⁶⁶, which fulfils the role of RNase E in *B. subtilis*¹⁶⁷. Correspondingly, the utility of 5' hairpins for protection of mRNA has been demonstrated in clostridia; the introduction of 5' stem-loop sequences was confirmed to increase mRNA stability, reporter expression, and expression of the

genes *adhE1* and *adhE2* in both *C. acetobutylicum* and *C. beijerinckii*¹⁶⁸. This effect was much more pronounced during solventogenesis than during acidogenesis. Similarly, the introduction of a terminator hairpin in the 3' UTR can result in improved mRNA stability via inhibition of nuclease activity¹²². Although not fully explored, this principle has been demonstrated in solventogenic clostridia; the expression of a *cat* reporter gene in *C. acetobutylicum* was observed to increase by approximately 36% when the downstream *adc* terminator was replaced by a synthetic terminator, BBa_B1010, from the iGEM registry¹⁶⁸. Additionally, a terminator with activity in the reverse orientation prevents the formation of antisense transcripts which are known to reduce protein expression levels¹⁶⁹.

A completely opposite approach is to reduce mRNA stability by introducing RNase sites into the 3' UTR. This may be useful in the case of proteins which form inclusion bodies when overexpressed. In *E. coli*, sequences derived from the *cat* gene, which contains 28 RNase E sites, have been shown to reduce mRNA stability¹⁷⁰. When combined with two poorly soluble heterologous enzymes, the *cat*-derived 3' UTRs were shown to result in an increase in soluble protein, with concomitant increases in enzyme activity. The authors noted that this improvement could not be observed simply by using a weaker promoter; it was proposed that by limiting the stability of the mRNA, the number of proteins that could be produced from a single transcript was decreased, thus limiting the local concentrations of protein during translation.

Antisense RNAs and Riboswitches

Bacterial antisense RNAs (asRNAs) are transcripts with complementarity to another RNA. Several studies have used asRNAs were used to reduce native gene expression in *C. acetobutylicum*^{169,171} and in *C. pasteurianum*¹⁷², demonstrating that multiple routes to regulate gene expression are available in the solventogenic clostridia.

Riboregulators are another class of naturally-occurring and generally trans-activating asRNA elements that respond to a signal nucleic acid by Watson-Crick base pairing¹⁷³. They have defined sensor and effector domains and have been rationally designed to repress¹⁷⁴ and activate gene expression¹⁷⁵ in *E. coli*. Toehold switches (a synthetic cis-regulatory subgroup of riboregulators) that activate gene expression in the presence of

cognate RNAs rely on sequestering the RBS and start codon¹⁷⁶. Like riboswitches, riboregulators are known to be present in *Clostridium* genomes.

Riboswitches are RNA sequences that are able to bind to a soluble ligand, influencing the properties of the RNA. In nature, riboswitches typically contain a binding domain, or 'aptamer', and an 'expression platform' which mediates the effect¹⁷⁷. Binding leads to a change in conformation of the RNA, leading to formation of a secondary structure which can act as a terminator or an anti-terminator. While riboswitches are typically found in the 5' UTRs of mRNAs, some have been recently determined to control the expression of antisense RNAs or protein-sequestering small RNAs, while yet others have been shown to control access to recognition sequences such as RNase sites¹⁷⁸. The range of applications has been further developed by the creation of synthetic riboswitches. In bacteria, riboswitches have been developed that can influence translation initiation by inhibiting access to the RBS; ligand binding leads to a conformational change or even to self-cleavage, revealing the RBS and allowing translation¹⁷⁹. Riboswitches are found in all taxa and a number have been characterized in the solventogenic clostridia¹⁸⁰⁻¹⁸². However, riboswitches have not yet been used in the engineering of these organisms. As discussed earlier, only a limited range of inducible systems is available for the engineering of clostridia; the use of synthetic riboswitches could be a promising alternative for the creation of controlled promoters.

Replication

Plasmid origins of replication

Four replicons are in routine usage in solventogens: pCB102 (from *C. butyricum*), pBP1 (from *C. botulinum*), pCD6 (from *C. difficile*) and pIM13 (from *B. subtilis*)⁹⁶. The pIM13 replication origin is thought to replicate via rolling-circle replication¹⁸³ while there is evidence that pCD6 replicates in similar fashion to pIP404^{184,185} (which is either theta or strand-displacement). The replication mechanisms of pCB102 and pBP1 are unknown. In *C. acetobutylicum*, two other replicons that have been used are the pAM β 1 replicon and the pUB110 replicon¹⁸⁶; the pUB110 replicon was found to be somewhat more stable than pIM13, whereas the pAM β 1 replicon was highly unstable. The *C. beijerinckii* filamentous phage CAK1's origin of replication has been used in *C. beijerinckii* strains¹⁸⁷. Additionally, the development of a replicon specific for *C. saccharoperbutylacetonicum* N1-4 was reported in 2007¹⁸⁸, this replicon is identical to the origin of the endogenous plasmid from *C. saccharoperbutylacetonicum* N1-504¹⁵. A thermosensitive origin pWV01ts derived from *Lactococcus lactis cremoris*¹⁸⁹ has been shown to work in both *C. ljungdahlii* and *C. acetobutylicum*⁵¹. Segregation and transformation frequencies are available; however, more work is needed to determine copy number and compatibility groups.

Chromosomal integration

The integration of DNA into the genome, while not a 'part' in itself, is an important consideration for synthetic biology projects. Genomic integration has several advantages over plasmid-based expression strategies, including increased stability, removal of the requirement for antibiotic selection, and standardisation of copy number^{190–192}. However, there are other factors that must be considered when using chromosomal integration. One implication of the integration position is the copy number effect – genes closer to the origin have a higher copy number than ones near the terminus in exponentially dividing cells due to the mechanism of DNA replication. There is contradictory evidence as to the effects of chromosome location and levels of expression^{193,194,195}. The cause of some of the apparent transcriptional insulation of chromosomal genes (and thus promoter independence) was found to be due to the gene silencing activity of the HU-protein in *E. coli*¹⁹⁴, a protein that is also present in

Firmicutes¹⁹⁶. Genomic engineering in *E. coli* has been carried out extensively; lambda-Red recombineering¹⁹⁷ is well-established, and the utility of the newly developed CRISPR technique has been demonstrated in this species^{198,199}. However, in the solventogenic clostridia, until recently published examples of genomic integration had been still somewhat limited. An early enabling technology was ClosTron which adapted the L1.LtrB intron for use in *Clostridium*¹¹⁰. Another method for genomic integration is Allele-Coupled Exchange (ACE), as demonstrated in *C. acetobutylicum*^{24,25,190}. This is a homologous recombination-based method, where homology arms with different lengths are used to control the sequence of recombination events, and the second recombination leads to the generation of a selectable phenotype; currently, this involves either the truncation or repair of the *pyrE* gene or the activation of a promoterless antibiotic resistance gene by integration downstream of a strong chromosomal promoter such as *thlA*¹⁹⁰. Thus, one potential drawback to this method is that it only allows integration into a limited selection of loci. This drawback can be mitigated by the ability to carry out multiple rounds of iterative ACE, thereby making further genomic integrations into the same locus.

Nevertheless, many new developments have been made regarding the genetic manipulation of solventogenic clostridia. A variety of different allelic exchange-based strategies have been exemplified in *C. acetobutylicum*^{101,109,200,201} and *C. beijerinckii*²⁰²; while most of these studies have focused on the generation of in-frame deletions and subsequent complementation, Al-Hinai et al.¹⁰¹ demonstrate the integration of a heterologous gene through gene replacement. Furthermore, the generation of point mutations through recombineering has been demonstrated in *C. acetobutylicum*²⁰³, suggesting that the integration of DNA through this method may be feasible. Finally, mutant selection via CRISPR has been established in almost all of the main solventogenic species, with published examples in *C. beijerinckii*^{121,204}, *C. acetobutylicum*¹⁰⁴, *C. pasteurianum*²⁰⁵, and *C. saccharoperbutylacetonicum*²⁰⁶. Future developments are certain to expand the range of genomic modifications that can be made in these organisms.

Insulators

A biological part's adjacent sequences can have a profound effect on its behaviour compared to the sequence context in which it was characterized²⁰⁷. This poses a

fundamental challenge to the synthetic biology principle of part creation and characterization²⁰⁸. To counteract this issue in reproducibility genetic engineers have started utilizing a new class of parts called insulators²⁰⁹. Several strategies can be undertaken to insulate a part from its genetic context, and these can be split into two main categories: DNA-level insulators (such as simply using flanking buffer zones of sequence without a biological function or secondary structure) and RNA-level insulators. The latter includes post-transcriptional modification of RNA as well as the commonly used flanking double terminators¹²³ that prevent read-through transcription into synthetic gene constructs. Post-transcriptional insulators consist of inclusion of ribozyme-based insulators or using CRISPR-RNA-processing to decouple the 5' UTR from the coding sequence (CDS)^{208,210}.

Different parts require different types of insulation in order to achieve maximum reproducibility without compromising features such as strength.

Promoters used in synthetic biology projects are often minimal (-35 to -10 region) and without characterized transcriptional start sites; the inclusion of important functional elements such as the UP element has been previously recommended.²¹¹ Accordingly, the addition of an upstream and downstream insulating sequence has increased reproducibility in different genetic contexts²¹². It is worth noting that the strong *Clostridium* promoters that are in widespread use in the *Clostridium* community, *fdx* and *thlA*, are 200 and 150bp respectively. Even though this recombinant *thlA* is longer than a usual minimal promoter, a long 5'UTR contributes to this part's length (recombinant *thlA* is 59bp from its 5' end to the transcriptional start site), while the recombinant 5' end is slightly truncated to exclude a Rex NADH-dependent regulator binding site (as mentioned before). A study that reported the use of a minimal *thlA* promoter in *C. acetobutylicum*, *miniPthl* which has a truncated 5' UTR, did not test activity variation in different genetic contexts⁹¹.

Work in *E. coli* has demonstrated the utility of 'bicistronic RBS', where a leader peptide is translationally coupled with the CDS of interest, in improving reliability and context independence (downstream gene sequence) when a particular 5' UTR is combined with a new coding sequence²¹². An upstream RBS and start codon initiate translation of the leader peptide, the stop codon of which overlaps with the start of the downstream CDS, while the latter's RBS is positioned with the leader peptide. The ribosomes translating the leader peptide unfold the 5' UTR (of the downstream CDS), preventing it from

forming secondary structures with the mRNA of the downstream CDS and thus influencing translation. The same study employed standard transcriptional start site (“+1 promoter”): 5’ UTR junctions to minimize (or insulate against) unforeseen effects of combining promoters with new 5’ UTRs.

As mentioned above, the observation of ribosomal repression of transcription termination also necessitates the more widespread use of ‘distance’ insulators of a sequence without emergent function and secondary structure to separate the stop codon and the stem-loop hairpin¹²⁶. Such strategies are yet to be implemented in *Clostridium* engineering projects.

Discussion

The development of biorenewables as an alternative to petroleum-derived commodity chemicals and fuels has resulted in the emergence of new markets²¹³. Increasing productivity, broadening the range of feedstocks, improving tolerance to solvents and by-products are all existing challenges to achieving higher sustainability and ensuring the economic viability of *Clostridium*-derived biorenewables²¹⁴. *Clostridium* species remain important hosts for the biological production of solvents and their further development relies on the adaptation of novel methodologies such as synthetic biology and metabolic engineering. Much progress has been made in the latter with several projects improving industrially relevant strains; however, the availability of biological parts with known behaviour is one of the limiting factors for the rate and scale of research. The reasons for our anticipation are twofold: first, the physical availability of biological parts streamlines assembly of genetic pathways, and second, knowledge about parts’ behaviour is crucial in predicting and analyzing the behaviour of pathways and genetic circuits. We have shown in this review that there are multiple areas where improved knowledge of parts would provide a better toolkit for synthetic biologists using these organisms and consequently improve the rate at which industrial biotechnology and bioenergy process development can improve.

References

- 902 (1) Breed, R. S., Murray, E. G. D., and Smith, N. R. (1957) Bergey's manual of
903 determinative bacteriology 7th ed. Williams & Wilkins Co., Baltimore.
- 904 (2) Jones, D. T., and Woods, D. R. (1986) Acetone-butanol fermentation revisited.
905 *Microbiol. Rev.* 50, 484–524.
- 906 (3) Gheshlaghi, R., Scharer, J. M., Moo-Young, M., and Chou, C. P. (2009) Metabolic
907 pathways of clostridia for producing butanol. *Biotechnol. Adv.* 27, 764–781.
- 908 (4) Keis, S., Bennett, C. F., Ward, V. K., and Jones, D. T. (1995) Taxonomy and phylogeny
909 of industrial solvent-producing clostridia. *Int. J. Syst. Bacteriol.* 45, 693–705.
- 910 (5) Keis, S., Shaheen, R., and Jones, D. T. (2001) Emended descriptions of *Clostridium*
911 *acetobutylicum* and *Clostridium beijerinckii*, and descriptions of *Clostridium*
912 *saccharoperbutylacetonicum* sp. nov. and *Clostridium saccharobutylicum* sp. nov. *Int. J.*
913 *Syst. Evol. Microbiol.* 51, 2095–2103.
- 914 (6) Hellinger, E. (1947) *Clostridium aurantibutyricum* (n.sp.): A Pink Butyric Acid
915 *Clostridium*. *J. Gen. Microbiol.* 1, 203–210.
- 916 (7) Holt, R. A., Cairns, A. J., and Morris, J. G. (1988) Production of butanol by *Clostridium*
917 *puniceum* in batch and continuous culture. *Appl. Microbiol. Biotechnol.* 27, 319–324.
- 918 (8) Abd-Alla, M. H., Zohri, A. N. A., El-Enany, A. W. E., and Ali, S. M. (2015) Acetone-
919 butanol-ethanol production from substandard and surplus dates by Egyptian native
920 *Clostridium* strains. *Anaerobe* 32, 77–86.
- 921 (9) Sarchami, T., Munch, G., Johnson, E., Kießlich, S., and Rehmann, L. (2016) A Review of
922 Process-Design Challenges for Industrial Fermentation of Butanol from Crude Glycerol
923 by Non-Biphasic *Clostridium pasteurianum*. *Fermentation* 2, 13.
- 924 (10) Tashiro, Y., Yoshida, T., Noguchi, T., and Sonomoto, K. (2013) Recent advances and
925 future prospects for increased butanol production by acetone-butanol-ethanol
926 fermentation. *Eng. Life Sci.* 13, 432–445.
- 927 (11) Tamura, K., and Nei, M. (1993) Estimation of the number of nucleotide substitutions
928 in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.*
929 10, 512–26.
- 930 (12) Felsenstein, J. (1985) Confidence Limits on Phylogenies: An Approach Using the
931 Bootstrap. *Evolution (N. Y.)*. 39, 783–791.
- 932 (13) Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011)
933 MEGA5: molecular evolutionary genetics analysis using maximum likelihood,
934 evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–9.

935 (14) ABAS. (2015) Classification of prokaryotes (bacteria and archaea) into risk groups
936 (TRBA 466). *Bundesarbeitsblatt*. <http://www.baua.de> 19.

937 (15) Poehlein, A., Solano, J. D. M., Flitsch, S. K., Krabben, P., Winzer, K., Reid, S. J., Jones, D.
938 T., Green, E., Minton, N. P., Daniel, R., and Dürre, P. (2017) Microbial solvent formation
939 revisited by comparative genome analysis. *Biotechnol. Biofuels* 10, 58.

940 (16) Udaondo, Z., Duque, E., and Ramos, J. L. (2017) The pangenome of the genus
941 *Clostridium*. *Environ. Microbiol.* 19, 2588–2603.

942 (17) Shaheen, R., Shirley, M., and Jones, D. T. (2000) Comparative fermentation studies of
943 industrial strains belonging to four species of solvent-producing clostridia. *J. Mol.*
944 *Microbiol. Biotechnol.* 2, 115–124.

945 (18) Jang, Y.-S., Lee, J. Y., Lee, J., Park, J. H., Im, J. A., Eom, M.-H., Lee, J., Lee, S.-H., Song, H.,
946 Cho, J.-H., Seung, D. Y., and Lee, S. Y. (2012) Enhanced butanol production obtained by
947 reinforcing the direct butanol-forming route in *Clostridium acetobutylicum*. *MBio* 3,
948 e00314-12.

949 (19) Pyne, M. E., Sokolenko, S., Liu, X., Srirangan, K., Bruder, M. R., Aucoin, M. G., Moo-
950 Young, M., Chung, D. A., and Chou, C. P. (2016) Disruption of the reductive 1,3-
951 propanediol pathway triggers production of 1,2-propanediol for sustained glycerol
952 fermentation by *Clostridium pasteurianum*. *Appl. Environ. Microbiol.* 82, 5375–5388.

953 (20) Xiao, H., Li, Z., Jiang, Y., Yang, Y., Jiang, W., Gu, Y., and Yang, S. (2012) Metabolic
954 engineering of D-xylose pathway in *Clostridium beijerinckii* to optimize solvent
955 production from xylose mother liquid. *Metab. Eng.* 14, 569–578.

956 (21) Xiao, H., Gu, Y., Ning, Y., Yang, Y., Mitchell, W. J., Jiang, W., and Yang, S. (2011)
957 Confirmation and elimination of xylose metabolism bottlenecks in glucose
958 phosphoenolpyruvate-dependent phosphotransferase system-deficient *Clostridium*
959 *acetobutylicum* for simultaneous utilization of glucose, xylose, and arabinose. *Appl.*
960 *Environ. Microbiol.* 77, 7886–7895.

961 (22) Mingardon, F., Chanal, A., Tardif, C., and Fierobe, H. P. (2011) The issue of secretion
962 in heterologous expression of *Clostridium cellulolyticum* cellulase-encoding genes in
963 *Clostridium acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* 77, 2831–2838.

964 (23) Chanal, A., Mingardon, F., Bauzan, M., Tardif, C., and Fierobe, H. P. (2011) Scaffoldin
965 modules serving as “cargo” domains to promote the secretion of heterologous
966 cellulosomal cellulases by *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 77, 6277–
967 6280.

968 (24) Kovács, K., Willson, B. J., Schwarz, K., Heap, J. T., Jackson, A., Bolam, D. N. N., Winzer,
969 K., and Minton, N. P. (2013) Secretion and assembly of functional mini-cellulosomes
970 from synthetic chromosomal operons in *Clostridium acetobutylicum* ATCC 824.
971 *Biotechnol. Biofuels* 6, 117.

972 (25) Willson, B. J., Kovács, K., Wilding-Steele, T., Markus, R., Winzer, K., and Minton, N. P.
973 (2016) Production of a functional cell wall-anchored minicellulosome by recombinant
974 *Clostridium acetobutylicum* ATCC 824. *Biotechnol. Biofuels* 9, 109.

975 (26) Collas, F., Kuit, W., Clément, B., Marchal, R., López-Contreras, A. M., and Monot, F.
976 (2012) Simultaneous production of isopropanol, butanol, ethanol and 2,3-butanediol by
977 *Clostridium acetobutylicum* ATCC 824 engineered strains. *AMB Express* 2, 45.

978 (27) González-Pajuelo, M., Meynial-Salles, I., Mendes, F., Andrade, J. C., Vasconcelos, I.,
979 and Soucaille, P. (2005) Metabolic engineering of *Clostridium acetobutylicum* for the
980 industrial production of 1,3-propanediol from glycerol. *Metab. Eng.* 7, 329–336.

981 (28) Cai, X., and Bennett, G. N. (2011) Improving the *Clostridium acetobutylicum* butanol
982 fermentation by engineering the strain for co-production of riboflavin. *J. Ind. Microbiol.*
983 *Biotechnol.* 38, 1013–1025.

984 (29) Quinn, J. Y., Cox, R. S., Adler, A., Beal, J., Bhatia, S., Cai, Y., Chen, J., Clancy, K.,
985 Galdicki, M., Hillson, N. J., Le Novère, N., Maheshwari, A. J., McLaughlin, J. A., Myers, C. J.,
986 Umesh, P., Pocock, M., Rodriguez, C., Soldatova, L., Stan, G. B. V., Swainston, N., Wipat, A.,
987 and Sauro, H. M. (2015) SBOL Visual: A Graphical Language for Genetic Designs. *PLoS*
988 *Biol.* 13, e1002310.

989 (30) Mermelstein, L. D., and Papoutsakis, E. T. (1993) *In vivo* methylation in *Escherichia*
990 *coli* by the *Bacillus subtilis* phage ϕ 3T I methyltransferase to protect plasmids from
991 restriction upon transformation of *Clostridium acetobutylicum* ATCC 824. *Appl. Environ.*
992 *Microbiol.* 59, 1077–1081.

993 (31) Nölling, J., Breton, G., Omelchenko, M. V, Makarova, K. S., Zeng, Q., Gibson, G., Hong
994 Mei Lee, Dubois, J., Qiu, D., Hitti, J., Aldredge, T., Ayers, M., Bashirzadeh, R., Bochner, H.,
995 Boivin, M., Bross, S., Bush, D., Butler, C., Caron, A., Caruso, A., Cook, R., Daggett, P.,
996 Deloughery, C., Egan, J., Ellston, D., Engelstein, M., Ezedi, J., Gilbert, K., Goyal, A., Guerin, J.,
997 Ho, T., Holtham, K., Joseph, P., Keagle, P., Kozlovsky, J., LaPlante, M., LeBlanc, G., Lumm,
998 W., Majeski, A., McDougall, S., Mank, P., Mao, J. I., Nocco, D., Patwell, D., Phillips, J., Pothier,
999 B., Prabhakar, S., Richterich, P., Rice, P., Rosetti, D., Rossetti, M., Rubenfield, M., Sachdeva,
1000 M., Snell, P., Spadafora, R., Spitzer, L., Shimer, G., Thomann, H. U., Vicaire, R., Wall, K.,

1001 Wang, Y., Weinstock, K., Lai Peng Wong, Wonsey, A., Xu, Q., Zhang, L., Wolf, Y. I., Tatusov,
1002 R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., Daly, M. J., Bennett, G. N., Koonin, E. V,
1003 and Smith, D. R. (2001) Genome sequence and comparative analysis of the solvent-
1004 producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* 183, 4823–4838.

1005 (32) Oultram, J. D., Loughlin, M., Swinfield, T.-J. J., Brehm, J. K., Thompson, D. E., and
1006 Minton, N. P. (1988) Introduction of plasmids into whole cells of *Clostridium*
1007 *acetobutylicum* by electroporation. *FEMS Microbiol. Lett.* 56, 83–88.

1008 (33) Pyne, M. E., Moo-Young, M., Chung, D. A., and Chou, C. P. (2013) Development of an
1009 electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*.
1010 *Biotechnol. Biofuels* 6, 50.

1011 (34) Lesiak, J. M., Liebl, W., and Ehrenreich, A. (2014) Development of an *in vivo*
1012 methylation system for the solventogen *Clostridium saccharobutylicum* NCP 262 and
1013 analysis of two endonuclease mutants. *J. Biotechnol.* 188, 97–99.

1014 (35) Herman, N. A., Li, J., Bedi, R., Turchi, B., Liu, X., Miller, M. J., and Zhang, W. (2017)
1015 Development of a High-Efficiency Transformation Method and Implementation of
1016 Rational Metabolic Engineering for the Industrial Butanol Hyperproducer *Clostridium*
1017 *saccharoperbutylacetonicum* Strain N1-4. *Appl. Environ. Microbiol.* 83, AEM.02942-16.

1018 (36) Grosse-Honebrink, A., Schwarz, K. M., Wang, H., Minton, N. P., and Zhang, Y. (2017)
1019 Improving gene transfer in *Clostridium pasteurianum* through the isolation of rare
1020 hypertransformable variants. *Anaerobe* 48, 203–205.

1021 (37) Kolek, J., Sedlar, K., Provaznik, I., and Patakova, P. (2016) Dam and Dcm
1022 methylations prevent gene transfer into *Clostridium pasteurianum* NRRL B-598:
1023 development of methods for electrotransformation, conjugation, and sonoporation.
1024 *Biotechnol. Biofuels* 9, 14.

1025 (38) Oh, Y. H., Eom, G. T., Kang, K. H., Choi, J. W., Song, B. K., Lee, S. H., and Park, S. J.
1026 (2015) Optimized Transformation of Newly Constructed *Escherichia coli*-Clostridia
1027 Shuttle Vectors into *Clostridium beijerinckii*. *Appl. Biochem. Biotechnol.* 177, 226–236.

1028 (39) Pyne, M. E., Bruder, M., Moo-Young, M., Chung, D. A., and Chou, C. P. (2014)
1029 Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*.
1030 *Biotechnol. Adv.* 32, 623–641.

1031 (40) Minton, N. P., Ehsaan, M., Humphreys, C. M., Little, G. T., Baker, J., Henstra, A. M.,
1032 Liew, F., Kelly, M. L., Sheng, L., Schwarz, K., and Zhang, Y. (2016) A roadmap for gene
1033 system development in *Clostridium*. *Anaerobe* 41, 104–112.

1034 (41) Shaner, N. C., Steinbach, P. A., and Tsien, R. Y. (2005) A guide to choosing
1035 fluorescent proteins. *Nat. Methods* 2, 905–909.

1036 (42) Drepper, T., Eggert, T., Circolone, F., Heck, A., Krauß, U., Guterl, J.-K., Wendorff, M.,
1037 Losi, A., Gärtner, W., and Jaeger, K.-E. (2007) Reporter proteins for *in vivo* fluorescence
1038 without oxygen. *Nat. Biotechnol.* 25, 443–445.

1039 (43) Jean, D., Briolat, V., and Reyssset, G. (2004) Oxidative stress response in *Clostridium*
1040 *perfringens*. *Microbiology* 150, 1649–1659.

1041 (44) Hartman, A. H., Liu, H., and Melville, S. B. (2011) Construction and Characterization
1042 of a Lactose-Inducible Promoter System for Controlled Gene Expression in *Clostridium*
1043 *perfringens*. *Appl. Environ. Microbiol.* 77, 471–478.

1044 (45) Ransom, E. M., Williams, K. B., Weiss, D. S., and Ellermeier, C. D. (2014)
1045 Identification and characterization of a gene cluster required for proper rod shape, cell
1046 division, and pathogenesis in *Clostridium difficile*. *J. Bacteriol.* 196, 2290–2300.

1047 (46) Ransom, E. M., Ellermeier, C. D., and Weiss, D. S. (2015) Use of mCherry red
1048 fluorescent protein for studies of protein localization and gene expression in *Clostridium*
1049 *difficile*. *Appl. Environ. Microbiol.* 81, 1652–1660.

1050 (47) Ransom, E. M., Weiss, D. S., and Ellermeier, C. D. (2016) Use of mCherryOpt
1051 fluorescent protein in *Clostridium difficile*. *Methods Mol. Biol.* 1476, 53–67.

1052 (48) Schulz, F. (2013) Fluoreszenzproteine in *Clostridium acetobutylicum* - Ein neues *in*
1053 *vivo* Reportersystem. Universität Rostock.

1054 (49) Li, Y., Xu, T., Tschaplinski, T. J., Engle, N. L., Yang, Y., Graham, D. E., He, Z., and Zhou, J.
1055 (2014) Improvement of cellulose catabolism in *Clostridium cellulolyticum* by sporulation
1056 abolishment and carbon alleviation. *Biotechnol. Biofuels* 7, 25.

1057 (50) Bruder, M. R., Pyne, M. E., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016)
1058 Extending CRISPR-Cas9 technology from genome editing to transcriptional engineering
1059 in the genus *Clostridium*. *Appl. Environ. Microbiol.* 82, 6109–6119.

1060 (51) Molitor, B., Kirchner, K., Henrich, A. W., Schmitz, S., and Rosenbaum, M. A. (2016)
1061 Expanding the molecular toolkit for the homoacetogen *Clostridium ljungdahlii*. *Sci. Rep.*
1062 6, 31518.

1063 (52) Chapman, S., Faulkner, C., Kaiserli, E., Garcia-Mata, C., Savenkov, E. I., Roberts, A. G.,
1064 Oparka, K. J., and Christie, J. M. (2008) The photoreversible fluorescent protein iLOV
1065 outperforms GFP as a reporter of plant virus infection. *Proc. Natl. Acad. Sci.* 105, 20038–
1066 20043.

1067 (53) Christie, J. M., Hitomi, K., Arvai, A. S., Hartfield, K. a., Mettlen, M., Pratt, A. J., Tainer, J.
1068 a., and Getzoff, E. D. (2012) Structural tuning of the fluorescent protein iLOV for
1069 improved photostability. *J. Biol. Chem.* 287, 22295–22304.

1070 (54) Gawthorne, J. A., Reddick, L. E., Akpunarlieva, S. N., Beckham, K. S. H., Christie, J. M.,
1071 Alto, N. M., Gabrielsen, M., and Roe, A. J. (2012) Express Your LOV: An Engineered
1072 Flavoprotein as a Reporter for Protein Expression and Purification. *PLoS One* 7, e52962.

1073 (55) Buckley, A. M., Jukes, C., Candlish, D., Irvine, J. J., Spencer, J., Fagan, R. P., Roe, A. J.,
1074 Christie, J. M., Fairweather, N. F., and Douce, G. R. (2016) Lighting Up *Clostridium difficile*:
1075 Reporting Gene Expression Using Fluorescent Lov Domains. *Sci. Rep.* 6, 23463.

1076 (56) Khrenova, M. G., Nemukhin, A. V., and Domratcheva, T. (2015) Theoretical
1077 characterization of the flavin-based fluorescent protein iLOV and its Q489K mutant. *J.*
1078 *Phys. Chem. B* 119, 5176–5183.

1079 (57) Davari, M. D., Kopka, B., Wingen, M., Bocola, M., Drepper, T., Jaeger, K. E.,
1080 Schwaneberg, U., and Krauss, U. (2016) Photophysics of the LOV-Based Fluorescent
1081 Protein Variant iLOV-Q489K Determined by Simulation and Experiment. *J. Phys. Chem. B*
1082 120, 3344–3352.

1083 (58) Bullifent, H. L., Moir, A., and Titball, R. W. (1995) The construction of a reporter
1084 system and use for the investigation of *Clostridium perfringens* gene expression. *FEMS*
1085 *Microbiol. Lett.* 131, 99–105.

1086 (59) Zhang, Y., Grosse-Honebrink, A., and Minton, N. P. (2015) A universal mariner
1087 transposon system for forward genetic studies in the genus *Clostridium*. *PLoS One* 10,
1088 e0122411.

1089 (60) Shaw, W. V. (1975) Chloramphenicol acetyltransferase from chloramphenicol-
1090 resistant bacteria. *Methods Enzymol.* 43, 737–755.

1091 (61) Roberts, M. C., Swenson, C. D., Owens, L. M., and Smith, A. L. (1980) Characterization
1092 of chloramphenicol-resistant *Haemophilus influenzae*. *Antimicrob. Agents Chemother.* 18,
1093 610–615.

1094 (62) Tummala, S. B., Welker, N. E., and Papoutsakis, E. T. (1999) Development and
1095 characterization of a gene expression reporter system for *Clostridium acetobutylicum*
1096 ATCC 824. *Appl. Environ. Microbiol.* 65, 3793–3799.

1097 (63) Girbal, L., Mortier-Barrière, I., Raynaud, F., Rouanet, C., Croux, C., and Soucaille, P.
1098 (2003) Development of a sensitive gene expression reporter system and an inducible
1099 promoter-repressor system for *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 69,

1100 4985–4988.

1101 (64) Moberg, L. J. (1985) Fluorogenic assay for rapid detection of *Escherichia coli* in food.

1102 *Appl. Environ. Microbiol.* 50, 1383–1387.

1103 (65) Sabathé, F., Croux, C., Cornillot, E., and Soucaille, P. (2002) *amyP*, a reporter gene to

1104 study strain degeneration in *Clostridium acetobutylicum* ATCC 824. *FEMS Microbiol. Lett.*

1105 210, 93–98.

1106 (66) Oliveira Paiva, A. M., Friggen, A. H., Hossein-Javaheri, S., and Smits, W. K. (2016) The

1107 Signal Sequence of the Abundant Extracellular Metalloprotease PPEP-1 Can Be Used to

1108 Secrete Synthetic Reporter Proteins in *Clostridium difficile*. *ACS Synth. Biol.* 5, 1376–

1109 1382.

1110 (67) Quixley, K. W., and Reid, S. J. (2000) Construction of a reporter gene vector for

1111 *Clostridium beijerinckii* using a *Clostridium* endoglucanase gene. *J. Mol. Microbiol.*

1112 *Biotechnol.* 2, 53–7.

1113 (68) Oh, Y. H., Eom, G. T., Kang, K. H., Joo, J. C., Jang, Y. A., Choi, J. W., Song, B. K., Lee, S. H.,

1114 and Park, S. J. (2016) Construction of heterologous gene expression cassettes for the

1115 development of recombinant *Clostridium beijerinckii*. *Bioprocess Biosyst. Eng.* 39, 555–

1116 563.

1117 (69) Gilham, D., and Lehner, R. (2005) Techniques to measure lipase and esterase

1118 activity *in vitro*. *Methods* 36, 139–147.

1119 (70) Hassan, E. A., Abd-Alla, M. H., Bagy, M. M. K., and Morsy, F. M. (2015) In situ

1120 hydrogen, acetone, butanol, ethanol and microdiesel production by *Clostridium*

1121 *acetobutylicum* ATCC 824 from oleaginous fungal biomass. *Anaerobe* 34, 125–131.

1122 (71) Edwards, A. N., Pascual, R. A., Childress, K. O., Nawrocki, K. L., Woods, E. C., and

1123 McBride, S. M. (2015) An alkaline phosphatase reporter for use in *Clostridium difficile*.

1124 *Anaerobe* 32, 98–104.

1125 (72) Feustel, L., Nakotte, S., and Dürre, P. (2004) Characterization and development of

1126 two reporter gene systems for *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 70,

1127 798–803.

1128 (73) Potzkei, J., Kunze, M., Drepper, T., Gensch, T., Jaeger, K.-E., and Buechs, J. (2012)

1129 Real-time determination of intracellular oxygen in bacteria using a genetically encoded

1130 FRET-based biosensor. *BMC Biol.* 10, 28.

1131 (74) Martin, L., Che, A., and Endy, D. (2009) Gemini, a bifunctional enzymatic and

1132 fluorescent reporter of gene expression. *PLoS One* 4.

1133 (75) Haima, P., van Sinderen, D., Bron, S., and Venema, G. (1990) An improved β -
1134 galactosidase α -complementation system for molecular cloning in *Bacillus subtilis*. *Gene*
1135 93, 41–47.

1136 (76) Knipfer, N., Nooruddin, L., and Shrader, T. E. (1998) Development of an alpha-
1137 complementation system for mycobacterial promoter analysis. *Gene* 217, 69–75.

1138 (77) Volkov, I. Y., Lunina, N. A., Berezina, O. V., Velikodvorskaya, G. A., and Zverlov, V. V.
1139 (2005) *Thermoanaerobacter ethanolicus* gene cluster containing the α - and β -
1140 galactosidase genes *mela* and *lacA* and properties of recombinant LacA. *Mol. Biol.* 39,
1141 799–805.

1142 (78) Huber, R. E., Roth, N. J., and Bahl, H. (1996) Quaternary structure, Mg²⁺
1143 interactions, and some kinetic properties of the β -galactosidase from
1144 *Thermoanaerobacterium thermosulfurigenes* EM1. *J. Protein Chem.* 15, 621–629.

1145 (79) Browning, D. F., and Busby, S. J. W. (2004) The regulation of bacterial transcription
1146 initiation. *Nat. Rev. Microbiol.* 2, 57–65.

1147 (80) Helmann, J. D., and Chamberlin, M. J. (1988) Structure and function of bacterial
1148 sigma factors. *Annu. Rev. Biochem.* 57, 839–872.

1149 (81) Sauer, U., Santangelo, J. D., Treuner, A., Buchholz, M., and Dürre, P. (1995) Sigma
1150 factor and sporulation genes in *Clostridium*. *FEMS Microbiol. Rev.* 17, 331–340.

1151 (82) Estrem, S. T., Gaal, T., Ross, W., and Gourse, R. L. (1998) Identification of an UP
1152 element consensus sequence for bacterial promoters. *Biochemistry* 95, 9761–9766.

1153 (83) Kumar, A., Malloch, R. A., Fujita, N., Smillie, D. A., Ishihama, A., and Hayward, R. S.
1154 (1993) The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for
1155 initiation of transcription at an “extended minus 10” promoter. *J. Mol. Biol.* 232, 406–18.

1156 (84) Sinoquet, C., Demey, S., and Braun, F. (2008) Large-scale computational and
1157 statistical analyses of high transcription potentialities in 32 prokaryotic genomes.
1158 *Nucleic Acids Res.* 36, 3332–3340.

1159 (85) Katayama, S., Matsushita, O., Jung, C. M., Minami, J., and Okabe, A. (1999) Promoter
1160 upstream bent DNA activates the transcription of the *Clostridium perfringens*
1161 phospholipase C gene in a low temperature-dependent manner. *EMBO J.* 18, 3442–3450.

1162 (86) Graves, M. C., and Rabinowitz, J. C. (1986) *In vivo* and *in vitro* transcription of the
1163 *Clostridium pasteurianum* ferredoxin gene. Evidence for “extended” promoter elements
1164 in gram-positive organisms. *J. Biol. Chem.* 261, 11409–15.

1165 (87) Pyne, M. E. (2014, April) Development of genetic tools for metabolic engineering of

1166 *Clostridium pasteurianum*. University of Waterloo.

1167 (88) Wiesenborn, D. P., Rudolph, F. B., and Papoutsakis, E. T. (1988) Thiolase from

1168 *Clostridium acetobutylicum* ATCC 824 and Its Role in the Synthesis of Acids and Solvents.

1169 *Appl Env. Microbiol* 54, 2717–2722.

1170 (89) Stim-Herndon, K. P., Petersen, D. J., and Bennett, G. N. (1995) Characterization of an

1171 acetyl-CoA C-acetyltransferase (thiolase) gene from *Clostridium acetobutylicum* ATCC

1172 824. *Gene*, pp 81–85.

1173 (90) Bormann, S., Baer, Z. C., Sreekumar, S., Kuchenreuther, J. M., Dean Toste, F., Blanch,

1174 H. W., and Clark, D. S. (2014) Engineering *Clostridium acetobutylicum* for production of

1175 kerosene and diesel blendstock precursors. *Metab. Eng.* 25, 124–130.

1176 (91) Dong, H., Tao, W., Zhang, Y., and Li, Y. (2012) Development of an

1177 anhydrotetracycline-inducible gene expression system for solvent-producing

1178 *Clostridium acetobutylicum*: A useful tool for strain engineering. *Metab. Eng.* 14, 59–67.

1179 (92) Perret, S., Casalot, L., Fierobe, H. P., Tardif, C., Sabathe, F., Belaich, J. P., and Belaich,

1180 A. (2004) Production of Heterologous and Chimeric Scaffoldins by *Clostridium*

1181 *acetobutylicum* ATCC 824. *J. Bacteriol.* 186, 253–257.

1182 (93) Mingardon, F., Perret, S., Bélaïch, A., Tardif, C., Bélaïch, J. P., and Fierobe, H. P. (2005)

1183 Heterologous production, assembly, and secretion of a minicellulosome by *Clostridium*

1184 *acetobutylicum* ATCC 824. *Appl. Environ. Microbiol.* 71, 1215–1222.

1185 (94) Wietzke, M., and Bahl, H. (2012) The redox-sensing protein Rex, a transcriptional

1186 regulator of solventogenesis in *Clostridium acetobutylicum*. *Appl. Microbiol. Biotechnol.*

1187 96, 749–761.

1188 (95) Zhang, L., Nie, X., Ravcheev, D. A., Rodionov, D. A., Sheng, J., Gu, Y., Yang, S., Jiang, W.,

1189 and Yang, C. (2014) Redox-responsive repressor rex modulates alcohol production and

1190 oxidative stress tolerance in *Clostridium acetobutylicum*. *J. Bacteriol.* 196, 3949–3963.

1191 (96) Heap, J. T., Pennington, O. J., Cartman, S. T., and Minton, N. P. (2009) A modular

1192 system for *Clostridium* shuttle plasmids. *J. Microbiol. Methods* 78, 79–85.

1193 (97) Novichkov, P. S., Kazakov, A. E., Ravcheev, D. A., Leyn, S. A., Kovaleva, G. Y., Sutormin,

1194 R. A., Kazanov, M. D., Riehl, W., Arkin, A. P., Dubchak, I., and Rodionov, D. A. (2013)

1195 RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in

1196 bacteria. *BMC Genomics* 14, 745.

1197 (98) Wang, Y., Zhang, Z.-T., Seo, S.-O., Lynn, P., Lu, T., Jin, Y.-S., and Blaschek, H. P. (2016)

1198 Gene transcription repression in *Clostridium beijerinckii* using CRISPR-dCas9. *Biotechnol.*

1199 *Bioeng. 113*, 2739–2743.

1200 (99) Pyne, M. E., Moo-Young, M., Chung, D. A., and Chou, C. P. (2014) Expansion of the
 1201 genetic toolkit for metabolic engineering of *Clostridium pasteurianum*: chromosomal
 1202 gene disruption of the endogenous CpaAI restriction enzyme. *Biotechnol. Biofuels* 7, 163.

1203 (100) Wang, Y., Li, X., Mao, Y., and Blaschek, H. P. (2012) Genome-wide dynamic
 1204 transcriptional profiling in *Clostridium beijerinckii* NCIMB 8052 using single-nucleotide
 1205 resolution RNA-Seq. *BMC Genomics* 13, 102.

1206 (101) Al-Hinai, M. A., Fast, A. G., and Papoutsakis, E. T. (2012) Novel system for efficient
 1207 isolation of *Clostridium* double-crossover allelic exchange mutants enabling markerless
 1208 chromosomal gene deletions and DNA integration. *Appl. Environ. Microbiol.* 78, 8112–
 1209 8121.

1210 (102) Garcia, H. G., Sanchez, A., Boedicker, J. Q., Osborne, M., Gelles, J., Kondev, J., and
 1211 Phillips, R. (2012) Operator sequence alters gene expression independently of
 1212 transcription factor occupancy in bacteria. *Cell Rep.* 2, 150–161.

1213 (103) Zong, Y., Zhang, H. M., Lyu, C., Ji, X., Hou, J., Guo, X., Ouyang, Q., and Lou, C. (2017)
 1214 Insulated transcriptional elements enable precise design of genetic circuits. *Nat.*
 1215 *Commun.* 8, 52.

1216 (104) Li, Q., Chen, J., Minton, N. P., Zhang, Y., Wen, Z., Liu, J., Yang, H., Zeng, Z., Ren, X.,
 1217 Yang, J., Gu, Y., Jiang, W., Jiang, Y., and Yang, S. (2016) CRISPR-based genome editing and
 1218 expression control systems in *Clostridium acetobutylicum* and *Clostridium beijerinckii*.
 1219 *Biotechnol. J.* 11, 961–972.

1220 (105) Tabor, S., and Richardson, C. C. (1985) A bacteriophage T7 RNA
 1221 polymerase/promoter system for controlled exclusive expression of specific genes (T7
 1222 DNA polymerase/T7 gene 5 protein/proteolysis/ β -lactamase/rifampicin). *Biochemistry*
 1223 24, 1074–1078.

1224 (106) McLaughlin, J. R., Murray, C. L., and Rabinowitz, J. C. (1981) Unique features in the
 1225 ribosome binding site sequence of the gram-positive *Staphylococcus aureus* beta-
 1226 lactamase gene. *J. Biol. Chem.* 256, 11283–91.

1227 (107) Lamberte, L. E., Baniulyte, G., Singh, S. S., Stringer, A. M., Bonocora, R. P., Stracy, M.,
 1228 Kapanidis, A. N., Wade, J. T., and Grainger, D. C. (2017) Horizontally acquired AT-rich
 1229 genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nat. Microbiol.*
 1230 2, 16249.

1231 (108) Heap, J. T., Kuehne, S. A., Ehsaan, M., Cartman, S. T., Cooksley, C. M., Scott, J. C., and

1232 Minton, N. P. (2010) The ClosTron: Mutagenesis in *Clostridium* refined and streamlined.
1233 *J. Microbiol. Methods* 80, 49–55.

1234 (109) Ehsaan, M., Kuit, W., Zhang, Y., Cartman, S. T., Heap, J. T., Winzer, K., and Minton, N.
1235 P. (2016) Mutant generation by allelic exchange and genome resequencing of the
1236 biobutanol organism *Clostridium acetobutylicum* ATCC 824. *Biotechnol. Biofuels* 9, 4.

1237 (110) Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P., and Minton, N. P. (2007)
1238 The ClosTron: A universal gene knock-out system for the genus *Clostridium*. *J. Microbiol.*
1239 *Methods* 70, 452–464.

1240 (111) Geissendörfer, M., and Hillen, W. (1990) Regulated expression of heterologous
1241 genes in *Bacillus subtilis* using the Tn10 encoded *tet* regulatory elements. *Appl. Microbiol.*
1242 *Biotechnol.* 33, 657–63.

1243 (112) Fontaine, L., Meynial-Salles, I., Girbal, L., Yang, X., Croux, C., and Soucaille, P. (2002)
1244 Molecular characterization and transcriptional analysis of *adhE2*, the gene encoding the
1245 NADH-dependent aldehyde/alcohol dehydrogenase responsible for butanol production
1246 in alcohologenic cultures of *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* 184, 821–
1247 830.

1248 (113) Ciampi, M. S. (2006) Rho-dependent terminators and transcription termination.
1249 *Microbiology* 152, 2515–2528.

1250 (114) Liu, C. C., Qi, L., Lucks, J. B., Segall-Shapiro, T. H., Wang, D., Mutalik, V. K., and Arkin,
1251 A. P. (2012) An adaptor from translational to transcriptional control enables predictable
1252 assembly of complex regulation. *Nat. Methods* 9, 1088–1094.

1253 (115) Banerjee, S., Chalissery, J., Bandey, I., and Sen, R. (2006) Rho-dependent
1254 transcription termination: more questions than answers. *J. Microbiol.* 44, 11–22.

1255 (116) Gardner, P. P., Barquist, L., Bateman, A., Nawrocki, E. P., and Weinberg, Z. (2011)
1256 RNIE: Genome-wide prediction of bacterial intrinsic terminators. *Nucleic Acids Res.* 39,
1257 5845–5852.

1258 (117) Gusarov, I., and Nudler, E. (1999) The mechanism of intrinsic transcription
1259 termination. *Mol. Cell* 3, 495–504.

1260 (118) Sugimoto, N., Nakano, S. ichi, Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T.,
1261 Yoneyama, M., and Sasaki, M. (1995) Thermodynamic Parameters To Predict Stability of
1262 RNA/DNA Hybrid Duplexes. *Biochemistry* 34, 11211–11216.

1263 (119) Naville, M., and Gautheret, D. (2010) Premature terminator analysis sheds light on
1264 a hidden world of bacterial transcriptional attenuation. *Genome Biol.* 11, R97.

1265 (120) Ravagnani, A., Jennert, K. C. B., Steiner, E., Grünberg, R., Jefferies, J. R., Wilkinson, S.
1266 R., Young, D. I., Tidswell, E. C., Brown, D. P., Youngman, P., Gareth Morris, J., and Young,
1267 M. (2000) Spo0A directly controls the switch from acid to solvent production in solvent-
1268 forming clostridia. *Mol. Microbiol.* 37, 1172–1185.

1269 (121) Wang, Y., Zhang, Z. T., Seo, S. O., Lynn, P., Lu, T., Jin, Y. S., and Blaschek, H. P. (2016)
1270 Bacterial Genome Editing with CRISPR-Cas9: Deletion, Integration, Single Nucleotide
1271 Modification, and Desirable “Clean” Mutant Selection in *Clostridium beijerinckii* as an
1272 Example. *ACS Synth. Biol.* 5, 721–732.

1273 (122) Carrier, T. A., and Keasling, J. D. (1997) Controlling messenger RNA stability in
1274 bacteria: Strategies for engineering gene expression. *Biotechnol. Prog.* 13, 699–708.

1275 (123) Chen, Y.-J., Liu, P., Nielsen, A. A. K., Brophy, J. A. N., Clancy, K., Peterson, T., and
1276 Voigt, C. A. (2013) Characterization of 582 natural and synthetic terminators and
1277 quantification of their design constraints. *Nat. Methods* 10, 659–664.

1278 (124) Lin, M. T., Wang, C. Y., Xie, H. J., Cheung, C. H. Y., Hsieh, C. H., Juan, H. F., Chen, B.
1279 Sen, and Lin, C. (2016) Novel Utilization of Terminators in the Design of Biologically
1280 Adjustable Synthetic Filters. *ACS Synth. Biol.* 5, 365–374.

1281 (125) Abe, H., Abo, T., and Aiba, H. (1999) Regulation of intrinsic terminator by
1282 translation in *Escherichia coli*: transcription termination at a distance downstream.
1283 *Genes Cells* 4, 87–97.

1284 (126) Li, R., Zhang, Q., Li, J., and Shi, H. (2015) Effects of cooperation between translating
1285 ribosome and RNA polymerase on termination efficiency of the Rho-independent
1286 terminator. *Nucleic Acids Res.* 44, 2554–2563.

1287 (127) Lesnik, E. A., Sampath, R., Levene, H. B., Henderson, T. J., McNeil, J. A., and Ecker, D.
1288 J. (2001) Prediction of rho-independent transcriptional terminators in *Escherichia coli*.
1289 *Nucleic Acids Res.* 29, 3583–94.

1290 (128) Kingsford, C. L., Ayanbule, K., and Salzberg, S. L. (2007) Rapid, accurate,
1291 computational discovery of Rho-independent transcription terminators illuminates their
1292 relationship to DNA uptake. *Genome Biol.* 8, R22.

1293 (129) Mitra, A., Angamuthu, K., Jayashree, H. V., and Nagaraja, V. (2009) Occurrence,
1294 divergence and evolution of intrinsic terminators across Eubacteria. *Genomics* 94, 110–
1295 116.

1296 (130) Mitra, A., Kesarwani, A. K., Pal, D., and Nagaraja, V. (2011) WebGeSTer DB-A
1297 transcription terminator database. *Nucleic Acids Res.* 39, D129-35.

1298 (131) Wright, J. J., Kumar, A., and Hayward, R. S. (1992) Hypersymmetry in a
1299 transcriptional terminator of *Escherichia coli* confers increased efficiency as well as
1300 bidirectionality. *EMBO J.* 11, 1957–64.

1301 (132) Yang, M. Te, Scott, H. B., and Gardner, J. F. (1995) Transcription termination at the
1302 *thr* attenuator: Evidence that the adenine residues upstream of the stem and loop
1303 structure are not required for termination. *J. Biol. Chem.* 270, 23330–23336.

1304 (133) Paredes, C. J., Rigoutsos, I., and Papoutsakis, T. (2004) Transcriptional
1305 organization of the *Clostridium acetobutylicum* genome. *Nucleic Acids Res.* 32, 1973–
1306 1981.

1307 (134) Shultzaberger, R. K., Bucheimer, R. E., Rudd, K. E., and Schneider, T. D. (2001)
1308 Anatomy of *Escherichia coli* ribosome binding sites. *J. Mol. Biol.* 313, 215–228.

1309 (135) Yang, Y., Lang, N., Yang, G., Yang, S., Jiang, W., and Gu, Y. (2016) Improving the
1310 performance of solventogenic clostridia by reinforcing the biotin synthetic pathway.
1311 *Metab. Eng.* 35, 121–128.

1312 (136) Yang, G., Jia, D., Jin, L., Jiang, Y., Wang, Y., Jiang, W., and Gu, Y. (2017) Rapid
1313 Generation of Universal Synthetic Promoters for Controlled Gene Expression in Both
1314 Gas-Fermenting and Saccharolytic *Clostridium* Species. *ACS Synth. Biol.* 6, 1672–1678.

1315 (137) Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994) Determination of the optimal
1316 aligned spacing between the Shine - Dalgarno sequence and the translation initiation
1317 codon of *Escherichia coli* mRNAs. *Nucleic Acids Res.* 22, 4953–4957.

1318 (138) Hager, P. W., and Rabinowitz, J. C. (1985) Translational specificity in *Bacillus*
1319 *subtilis*. *Mol. Biol. bacilli* 2, 1–32.

1320 (139) Band, L., and Henner, D. J. (1984) *Bacillus subtilis* requires a “stringent” Shine-
1321 Dalgarno region for gene expression. *DNA* 3, 17–21.

1322 (140) Abolbaghaei, A., Silke, J. R., and Xia, X. (2017) How Changes in Anti-SD Sequences
1323 Would Affect SD Sequences in *Escherichia coli* and *Bacillus subtilis*. *G3 (Bethesda)*. 7,
1324 1607–1615.

1325 (141) Dunn, J. J., and Studier, F. W. (1981) Nucleotide sequence from the genetic left end
1326 of bacteriophage T7 DNA to the beginning of gene 4. *J. Mol. Biol.* 148, 303–330.

1327 (142) Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., and
1328 Gold, L. (1992) Translation initiation in *Escherichia coli*: sequences within the ribosome-
1329 binding site. *Mol. Microbiol.* 6, 1219–1229.

1330 (143) Garnier, T., Canard, B., and Cole, S. T. (1991) Cloning, mapping, and molecular

1331 characterization of the rRNA operons of *Clostridium perfringens*. *J. Bacteriol.* 173, 5431–
1332 5438.

1333 (144) Ma, J., Campbell, A., and Karlin, S. (2002) Correlations between Shine-Dalgarno
1334 sequences and gene features such as predicted expression levels and operon structures.
1335 *J. Bacteriol.* 184, 5733–5745.

1336 (145) He, J., Sakaguchi, K., and Suzuki, T. (2012) Determination of the ribosome-binding
1337 sequence and spacer length between binding site and initiation codon for efficient
1338 protein expression in *Bifidobacterium longum* 105-A. *J. Biosci. Bioeng.* 113, 442–444.

1339 (146) Salis, H. M. (2011) The ribosome binding site calculator. *Methods Enzymol.* 498,
1340 19–42.

1341 (147) Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic
1342 ribosome binding sites to control protein expression. *Nat. Biotechnol.* 27, 946–950.

1343 (148) Reeve, B., Martinez-Klimova, E., De Jonghe, J., Leak, D. J., and Ellis, T. (2016) The
1344 *Geobacillus* Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synth. Biol.*
1345 5, 1342–1347.

1346 (149) Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular
1347 organisms. *Mol. Biol. Evol.* 2, 13–34.

1348 (150) Sharp, P. M., and Li, W. H. (1987) The codon adaptation index-a measure of
1349 directional synonymous codon usage bias, and its potential applications. *Nucleic Acids*
1350 *Res.* 15, 1281–1295.

1351 (151) Baca, A. M., and Hol, W. G. J. (2000) Overcoming codon bias: A method for high-
1352 level overexpression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*.
1353 *Int. J. Parasitol.* 30, 113–118.

1354 (152) Novy, R., Drott, D., Yaeger, K., and Mierendorf, R. (2001) Overcoming the codon
1355 bias of *E. coli* for enhanced protein expression. *Innovations* 12, 4–6.

1356 (153) Mondal, S. K., Kundu, S., Das, R., and Roy, S. (2016) Analysis of phylogeny and
1357 codon usage bias and relationship of GC content, amino acid composition with
1358 expression of the structural *nif* genes. *J. Biomol. Struct. Dyn.* 34, 1649–1666.

1359 (154) Musto, H., Romero, H., and Zavala, A. (2003) Translational selection is operative
1360 for synonymous codon usage in *Clostridium perfringens* and *Clostridium acetobutylicum*.
1361 *Microbiology* 149, 855–863.

1362 (155) Sastalla, I., Chim, K., Cheung, G. Y. C., Pomerantsev, A. P., and Leppla, S. H. (2009)
1363 Codon-optimized fluorescent proteins designed for expression in low-GC gram-positive

1364 bacteria. *Appl. Environ. Microbiol.* 75, 2099–2110.

1365 (156) Wright, F. (1990) The “effective number of codons” used in a gene. *Gene* 87, 23–29.

1366 (157) Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009) Coding-Sequence
1367 Determinants of Gene Expression in *Escherichia coli*. *Science* (80-.). 324, 255–258.

1368 (158) Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., and
1369 Gustafsson, C. (2009) Design parameters to control synthetic gene expression in
1370 *Escherichia coli*. *PLoS One* 4, e7002.

1371 (159) Xia, X. (2007) An improved implementation of codon adaptation index. *Evol.*
1372 *Bioinforma.* 3, 53–58.

1373 (160) Xia, X. (2015) A major controversy in codon-Anticodon adaptation resolved by a
1374 new codon usage index. *Genetics* 199, 573–579.

1375 (161) Lanza, A. M., Curran, K. A., Rey, L. G., and Alper, H. S. (2014) A condition-specific
1376 codon optimization approach for improved heterologous gene expression in
1377 *Saccharomyces cerevisiae*. *BMC Syst. Biol.* 8, 33.

1378 (162) Welch, M., Villalobos, A., Gustafsson, C., and Minshull, J. (2009) You’re one in a
1379 googol: optimizing genes for protein expression. *J. R. Soc. Interface* 6, S467–S476.

1380 (163) Plotkin, J. B., and Kudla, G. (2011) Synonymous but not the same: the causes and
1381 consequences of codon bias. *Nat. Rev. Genet.* 12, 32–42.

1382 (164) Carrier, T. A., and Keasling, J. D. (1999) Library of synthetic 5’ secondary
1383 structures to manipulate mRNA stability in *Escherichia coli*. *Biotechnol. Prog.* 15, 58–64.

1384 (165) Condon, C., and Putzer, H. (2002) The phylogenetic distribution of bacterial
1385 ribonucleases. *Nucleic Acids Res.* 30, 5339–5346.

1386 (166) Durand, S., Tomasini, A., Braun, F., Condon, C., and Romby, P. (2015) sRNA and
1387 mRNA turnover in Gram-positive bacteria. *FEMS Microbiol. Rev.* 39, 316–330.

1388 (167) Lehnik-Habrink, M., Newman, J., Rothe, F. M., Solovyova, A. S., Rodrigues, C.,
1389 Herzberg, C., Commichau, F. M., Lewis, R. J., and Stülke, J. (2011) RNase Y in *Bacillus*
1390 *subtilis*: A natively disordered protein that is the functional equivalent of RNase E from
1391 *Escherichia coli*. *J. Bacteriol.* 193, 5431–5441.

1392 (168) Lee, J., Jang, Y. S., Papoutsakis, E. T., and Lee, S. Y. (2016) Stable and enhanced gene
1393 expression in *Clostridium acetobutylicum* using synthetic untranslated regions with a
1394 stem-loop. *J. Biotechnol.* 230, 40–43.

1395 (169) Desai, R. P., and Papoutsakis, E. T. (1999) Antisense RNA strategies for metabolic
1396 engineering of *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 65, 936–945.

1397 (170) Song, J.-W., Woo, J.-M., Jung, G. Y., Bornscheuer, U. T., and Park, J.-B. (2016) 3'-UTR
1398 engineering to improve soluble expression and fine-tuning of activity of cascade
1399 enzymes in *Escherichia coli*. *Sci. Rep.* 6, 29406.

1400 (171) Tummala, S. B., Junne, S. G., and Papoutsakis, E. T. (2003) Antisense RNA
1401 downregulation of coenzyme A transferase combined with alcohol-aldehyde
1402 dehydrogenase overexpression leads to predominantly alcohologenic *Clostridium*
1403 *acetobutylicum* fermentations. *J. Bacteriol.* 185, 3644–3653.

1404 (172) Pyne, M., Moo-Young, M., Chung, D., and Chou, C. (2015) Antisense-RNA-Mediated
1405 Gene Downregulation in *Clostridium pasteurianum*. *Fermentation* 1, 113–126.

1406 (173) Storz, G., Altuvia, S., and Wassarman, K. M. (2005) AN ABUNDANCE OF RNA
1407 REGULATORS. *Annu. Rev. Biochem.* 74, 199–217.

1408 (174) Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B., and Arkin, A. P. (2012) Rationally
1409 designed families of orthogonal RNA regulators of translation. *Nat. Chem. Biol.* 8, 447–
1410 454.

1411 (175) Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., and Collins, J. J.
1412 (2004) Engineered riboregulators enable post-transcriptional control of gene
1413 expression. *Nat. Biotechnol.* 22, 841–847.

1414 (176) Green, A. A., Silver, P. A., Collins, J. J., and Yin, P. (2014) Toehold switches: De-novo-
1415 designed regulators of gene expression. *Cell* 159, 925–939.

1416 (177) Garst, A. D., Edwards, A. L., and Batey, R. T. (2011) Riboswitches: Structures and
1417 mechanisms. *Cold Spring Harb. Perspect. Biol.* 3, 1–13.

1418 (178) Mellin, J. R., and Cossart, P. (2015) Unexpected versatility in bacterial
1419 riboswitches. *Trends Genet.* 31, 150–156.

1420 (179) Wittmann, A., and Suess, B. (2012) Engineered riboswitches: Expanding
1421 researchers' toolbox with synthetic RNA regulators. *FEBS Lett.* 586, 2076–2083.

1422 (180) Saad, N. Y., Schiel, B., Brayé, M., Heap, J. T., Minton, N. P., Dürre, P., and Becker, H. D.
1423 (2012) Riboswitch (T-box)-mediated control of tRNA-dependent amidation in
1424 *Clostridium acetobutylicum* rationalizes gene and pathway redundancy for asparagine
1425 and asparaginyln-tRNA^{Asn} synthesis. *J. Biol. Chem.* 287, 20382–20394.

1426 (181) André, G., Even, S., Putzer, H., Burguière, P., Croux, C., Danchin, A., Martin-
1427 Verstraete, I., and Soutourina, O. (2008) S-box and T-box riboswitches and antisense
1428 RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res.*
1429 36, 5955–5969.

1430 (182) Kim, P. B., Nelson, J. W., and Breaker, R. R. (2015) An ancient riboswitch class in
 1431 bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol. Cell* 57, 317–
 1432 328.

1433 (183) Projan, S. J., Monod, M., Narayanan, C. S., and Dubnau, D. (1987) Replication
 1434 properties of pIM13, a naturally occurring plasmid found in *Bacillus subtilis*, and of its
 1435 close relative pE5, a plasmid native to *Staphylococcus aureus*. *J. Bacteriol.* 169, 5131–
 1436 5139.

1437 (184) Garnier, T., and Cole, S. T. (1988) Identification and Molecular Genetic Analysis of
 1438 Replication Functions of the Bacteriocinogenic Plasmid plP404 from *Clostridium*
 1439 *perfringens*. *Plasmid* 19, 151–160.

1440 (185) Purdy, D., O’Keeffe, T. A. T., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M.,
 1441 Ostrowski, A., and Minton, N. P. (2002) Conjugative transfer of clostridial shuttle vectors
 1442 from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction
 1443 barrier. *Mol. Microbiol.* 46, 439–452.

1444 (186) Lee, S. H., Kwon, M. A., Choi, S., Kim, S., Kim, J., Shin, Y. A., and Kim, K. H. (2015) A
 1445 new shuttle plasmid that stably replicates in *Clostridium acetobutylicum*. *J. Microbiol.*
 1446 *Biotechnol.* 25, 1702–1708.

1447 (187) Li, Y., Blaschek, H. P., and Tan, D. (2002) Molecular characterization and utilization
 1448 of the CAK1 filamentous viruslike particle derived from *Clostridium beijerinckii*. *J. Ind.*
 1449 *Microbiol. Biotechnol.* 28, 118–26.

1450 (188) Nakayama, S., Irie, R., Kosaka, T., Matsuura, K., Yoshino, S., and Furukawa, K.
 1451 (2007) New host-vector system in solvent-producing *Clostridium*
 1452 *saccharoperbutylacetonicum* strain N1-4. *J. Gen. Appl. Microbiol.* 53, 53–6.

1453 (189) Maguin, E., Duwat, P., Hege, T., Ehrlich, D., and Gruss, A. (1992) New
 1454 thermosensitive plasmid for gram-positive bacteria. *J. Bacteriol.* 174, 5633–5638.

1455 (190) Heap, J. T., Ehsaan, M., Cooksley, C. M., Ng, Y. K., Cartman, S. T., Winzer, K., and
 1456 Minton, N. P. (2012) Integration of DNA into bacterial chromosomes from plasmids
 1457 without a counter-selection marker. *Nucleic Acids Res.* 40, e59.

1458 (191) Gu, P., Yang, F., Su, T., Wang, Q., Liang, Q., and Qi, Q. (2015) A rapid and reliable
 1459 strategy for chromosomal integration of gene(s) with multiple copies. *Sci. Rep.* 5, 9684.

1460 (192) Kuhlman, T. E., and Cox, E. C. (2010) Site-specific chromosomal integration of
 1461 large synthetic constructs. *Nucleic Acids Res.* 38, e92.

1462 (193) Bryant, J. A., Sellars, L. E., Busby, S. J. W., and Lee, D. J. (2014) Chromosome

1463 position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Res.* 42, 11383–
1464 11392.

1465 (194) Berger, M., Gerganova, V., Berger, P., Rapiteanu, R., and Lisicovas, V. (2016) Genes
1466 on a Wire : The Nucleoid- Associated Protein HU Insulates Transcription Units in
1467 *Escherichia coli*. *Nat. Publ. Gr.* 6, 1–12.

1468 (195) Sauer, C., Syvertsson, S., Bohorquez, L. C., Cruz, R., Harwood, C. R., Van Rij, T., and
1469 Hamoen, L. W. (2016) Effect of Genome Position on Heterologous Gene Expression in
1470 *Bacillus subtilis*: An Unbiased Analysis. *ACS Synth. Biol.* 5, 942–947.

1471 (196) Anuchin, A. M., Goncharenko, A. V, Demidenok, O. I., and Kaprelyants, A. S. (2011)
1472 Histone-like proteins of bacteria (review). *Appl. Biochem. Microbiol.* 47, 580–585.

1473 (197) Sharan, S. K., Thomason, L. C., Kuznetsov, S. G., and Court, D. L. (2009)
1474 Recombineering: a homologous recombination-based method of genetic engineering.
1475 *Nat. Protoc.* 4, 206–223.

1476 (198) Li, Y., Lin, Z., Huang, C., Zhang, Y., Wang, Z., Tang, Y. jie, Chen, T., and Zhao, X.
1477 (2015) Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome
1478 editing. *Metab. Eng.* 31, 13–21.

1479 (199) Jiang, Y., Chen, B., Duan, C., Sun, B., Yang, J., and Yang, S. (2015) Multigene editing
1480 in the *Escherichia coli* genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* 81,
1481 2506–2514.

1482 (200) Croux, C., Nguyen, N.-P.-T., Lee, J., Raynaud, C., Saint-Prix, F., Gonzalez-Pajuelo, M.,
1483 Meynial-Salles, I., and Soucaille, P. (2016) Construction of a restriction-less, marker-less
1484 mutant useful for functional genomic and metabolic engineering of the biofuel producer
1485 *Clostridium acetobutylicum*. *Biotechnol. Biofuels* 9, 23.

1486 (201) Yoo, M., Croux, C., Meynial-Salles, I., and Soucaille, P. (2016) Elucidation of the
1487 roles of *adhE1* and *adhE2* in the primary metabolism of *Clostridium acetobutylicum* by
1488 combining in-frame gene deletion and a quantitative system-scale approach. *Biotechnol.*
1489 *Biofuels* 9, 92.

1490 (202) Zhang, N., Shao, L., Jiang, Y., Gu, Y., Li, Q., Liu, J., Jiang, W., and Yang, S. (2015) I-SceI-
1491 mediated scarless gene modification via allelic exchange in *Clostridium*. *J. Microbiol.*
1492 *Methods* 108, 49–60.

1493 (203) Dong, H., Tao, W., Gong, F., Li, Y., and Zhang, Y. (2014) A functional *recT* gene for
1494 recombineering of *Clostridium*. *J. Biotechnol.* 173, 65–67.

1495 (204) Wang, Y., Zhang, Z. T., Seo, S. O., Choi, K., Lu, T., Jin, Y. S., and Blaschek, H. P. (2015)

1496 Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9
 1497 system. *J. Biotechnol.* 200, 1–5.
 1498 (205) Pyne, M. E., Bruder, M. R., Moo-Young, M., Chung, D. A., and Chou, C. P. (2016)
 1499 Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient
 1500 markerless genome editing in *Clostridium*. *Sci. Rep.* 6, 25666.
 1501 (206) Wang, S., Dong, S., Wang, P., Tao, Y., and Wang, Y. (2017) Genome editing in
 1502 *Clostridium saccharoperbutylacetonicum* N1-4 with the CRISPR-Cas9 system. *Appl.*
 1503 *Environ. Microbiol.* 83.
 1504 (207) Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., Endy, D.,
 1505 and Church, G. M. (2013) Composability of regulatory sequences controlling
 1506 transcription and translation in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 110, 14024–14029.
 1507 (208) Lou, C., Stanton, B., Chen, Y.-J., Munskey, B., and Voigt, C. A. (2012) Ribozyme-based
 1508 insulator parts buffer synthetic circuits from genetic context. *Nat. Biotechnol.* 30, 1137–
 1509 1142.
 1510 (209) Davis, J. H., Rubin, A. J., and Sauer, R. T. (2011) Design, construction and
 1511 characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.* 39, 1131–
 1512 1141.
 1513 (210) Qi, L., Haurwitz, R. E., Shao, W., Doudna, J. A., and Arkin, A. P. (2012) RNA
 1514 processing enables predictable programming of gene expression. *Nat. Biotechnol.* 30,
 1515 1002–1006.
 1516 (211) Nielsen, A. A. K., Segall-Shapiro, T. H., and Voigt, C. A. (2013) Advances in genetic
 1517 circuit design: Novel biochemistries, deep part mining, and precision gene expression.
 1518 *Curr. Opin. Chem. Biol.* 17, 878–892.
 1519 (212) Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A.,
 1520 Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable
 1521 gene expression via standard transcription and translation initiation elements. *Nat.*
 1522 *Methods* 10, 354–360.
 1523 (213) Bardhan, S. K., Gupta, S., Gorman, M. E., and Haider, M. A. (2015) Biorenewable
 1524 chemicals: Feedstocks, technologies and the conflict with food production. *Renew.*
 1525 *Sustain. Energy Rev.* 51, 506–520.
 1526 (214) Papoutsakis, E. T. (2008) Engineering solventogenic clostridia. *Curr. Opin.*
 1527 *Biotechnol.* 19, 420–9.
 1528 (215) Lorenz, R., Bernhart, S. S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C.,

1529 Stadler, P. F. P., and Hofacker, I. I. L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*
1530 6, 26.

1531 (216) Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008) The
1532 Vienna RNA websuite. *Nucleic Acids Res.* 36, W70-4.

1533 (217) Darty, K., Denise, A., and Ponty, Y. (2009) VARNA: Interactive drawing and editing
1534 of the RNA secondary structure. *Bioinformatics* 25, 1974–1975.

1535 (218) Wilson, K. S., and Hippel, P. H. Von. (1995) Transcription termination at intrinsic
1536 terminators: The role of the RNA hairpin (*Escherichia coli*/RNA polymerase/rho-
1537 independent termination). *Biochemistry* 92, 8793–8797.

1538